GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 19:49:27 ; Search time 117 Seconds
(without alignments)
5507.858 Million cell updates/sec

Title: US-09-595-947E-1
Perfect score: 1460
Sequence: 1 gcaggtagcgagaggagcag.....agagtgacctaatccagtgt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	92	6.3	804	4	US-08-932-411A-19	Sequence 19, Appl
	2	23	1.6	738	4	US-08-722-570-13	Sequence 13, Appl
	3	23	1.6	738	4	US-08-932-411A-13	Sequence 13, Appl
	4	23	1.6	1333	1	US-08-910-973-21	Sequence 21, Appl
	5	23	1.6	1333	4	US-09-499-227-21	Sequence 21, Appl
	6	23	1.6	1385	4	US-08-932-411A-17	Sequence 17, Appl
	7	21	1.4	310	1	US-08-552-142A-12	Sequence 12, Appl
	8	21	1.4	1268	1	US-08-910-973-12	Sequence 12, Appl
	9	21	1.4	1268	4	US-09-499-227-12	Sequence 12, Appl
	10	21	1.4	1352	1	US-08-552-142A-10	Sequence 10, Appl
	11	21	1.4	1535	1	US-08-910-973-10	Sequence 10, Appl
	12	21	1.4	1535	4	US-09-499-227-10	Sequence 10, Appl
	13	21	1.4	1550	3	US-09-234-332-3	Sequence 3, Appli
c	14	19	1.3	50	3	US-08-358-627F-4	Sequence 4, Appli
c	15	19	1.3	50	3	US-08-793-044-11	Sequence 11, Appl
c	16	19	1.3	50	4	US-08-465-712C-4	Sequence 4, Appli
c	17	19	1.3	50	4	US-09-552-733-4	Sequence 4, Appli
c	18	18	1.2	480	1	US-08-438-123-16	Sequence 16, Appl
	19	18	1.2	1336	4	US-09-016-434-129	Sequence 129, App
	20	18	1.2	1462	1	US-08-552-142A-16	Sequence 16, Appl
	21	18	1.2	1951	1	US-08-910-973-16	Sequence 16, Appl
	22	18	1.2	1951	4	US-09-499-227-16	Sequence 16, Appl
	23	18	1.2	3446	4	US-09-620-312D-653	Sequence 653, App
	24	18	1.2	4066	4	US-09-367-750-1	Sequence 1, Appli
	25	18	1.2	4797	4	US-09-419-568F-25	Sequence 25, Appl
	26	18	1.2	4797	4	US-09-354-243B-25	Sequence 25, Appl
c	27	18	1.2	31728	4	US-09-453-702B-64	Sequence 64, Appl
	28	17	1.2	24	3	US-08-358-627F-6	Sequence 6, Appli
	29	17	1.2	24	4	US-08-465-712C-6	Sequence 6, Appli
	30	17	1.2	24	4	US-09-552-733-6	Sequence 6, Appli
c	31	17	1.2	24	4	US-09-687-731-2	Sequence 2, Appli
c	32	17	1.2	24	4	US-09-687-731-8	Sequence 8, Appli
c	33	17	1.2	24	4	US-09-687-731-10	Sequence 10, Appl
	34	17	1.2	510	4	US-09-252-991A-9622	Sequence 9622, Ap
	35	17	1.2	685	1	US-08-751-782-5	Sequence 5, Appli
	36	17	1.2	685	2	US-08-925-171-5	Sequence 5, Appli
c	37	17	1.2	846	4	US-09-252-991A-9635	Sequence 9635, Ap
	38	17	1.2	948	4	US-09-252-991A-14810	Sequence 14810, A
	39	17	1.2	954	4	US-09-252-991A-1904	Sequence 1904, Ap
	40	17	1.2	1062	4	US-09-252-991A-9532	Sequence 9532, Ap
c	41	17	1.2	1161	1	US-08-086-439C-1	Sequence 1, Appli
c	42	17	1.2	1161	1	US-08-434-877-1	Sequence 1, Appli
c	43	17	1.2	1238	4	US-09-183-861-75	Sequence 75, Appl
c	44	17	1.2	1238	4	US-09-022-765-75	Sequence 75, Appl
c	45	17	1.2	1238	4	US-09-551-974A-75	Sequence 75, Appl

ALIGNMENTS

RESULT 1
US-08-932-411A-19
; Sequence 19, Application US/08932411A
; Patent No. 6566496
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qiufu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,411A
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/772,009
; FILING DATE: 19-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/722,570
; FILING DATE: 19-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 804 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 160..801
US-08-932-411A-19

```
  Query Match             6.3%;  Score 92;  DB 4;  Length 804;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-37;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 522

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              ||||||||||||||||||||||||||||||||
Db        523 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 554


RESULT 2
US-08-722-570-13
; Sequence 13, Application US/08722570
; Patent No. 6555337
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  5365
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-13

  Query Match             1.6%;  Score 23;  DB 4;  Length 738;
  Best Local Similarity   100.0%;  Pred. No. 0.072;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        304 GAGCGCAACCGCATGCACAACCT 326


RESULT 3
US-08-932-411A-13
; Sequence 13, Application US/08932411A
; Patent No. 6566496
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr Hohbach Test Albritton & Herbert LLP
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/932,411A
;      FILING DATE:  15-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/772,009
;      FILING DATE:  19-DEC-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,570
;      FILING DATE:  19-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902-3/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..732
US-08-932-411A-13

  Query Match             1.6%;  Score 23;  DB 4;  Length 738;
  Best Local Similarity   100.0%;  Pred. No. 0.072;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        304 GAGCGCAACCGCATGCACAACCT 326


RESULT 4
US-08-910-973-21
; Sequence 21, Application US/08910973
; Patent No. 5795723
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectode
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 2, 2004, 15:49:38 ; Search time 3511 Seconds
(without alignments)
2493.495 Million cell updates/sec

Title: US-09-595-947E-10
Perfect score: 1127
Sequence: 1 MTPQPSGAPTVQVTRETERS.....LGATSSACLSPGSLAFSDFL 214

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09595947/runat_02022004_154933_8046/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09595947_@CGN_1_1_3508_@runat_02022004_154933_8046 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1127	100.0	1330	9	HSA133776	AJ133776 Homo sapi
2	1105	98.0	5340	9	AF234829	AF234829 Homo sapi
c 3	1099	97.5	165110	9	AL450311	AL450311 Human DNA
c 4	1089	96.6	173341	2	AC021954	AC021954 Homo sapi
5	849	75.3	1491	6	A91167	A91167 Sequence 1
6	849	75.3	1491	6	BD023626	BD023626 Polypepti
7	849	75.3	1491	10	RNRELAXT	Y10619 R.norvegicu
8	849	75.3	258815	2	AC127817	AC127817 Rattus no
9	819	72.7	861	6	AX698801	AX698801 Sequence
10	819	72.7	861	10	MMU76208	U76208 Mus musculu
11	819	72.7	1861	10	AF364300	AF364300 Mus muscu
12	819	72.7	5567	10	MMMATH4B	Y09167 M.musculus
c 13	819	72.7	138070	2	AC109783	AC109783 Mus muscu
14	819	72.7	215050	2	AC127417	AC127417 Mus muscu
c 15	523	46.4	170896	2	AC011010	AC011010 Homo sapi
16	388.5	34.5	790	5	GGA012659	AJ012659 Gallus ga
17	388.5	34.5	1880	5	AF303000	AF303000 Gallus ga
18	385.5	34.2	1074	5	GGA012660	AJ012660 Gallus ga
19	382	33.9	1299	5	AF123883	AF123883 Gallus ga
20	381.5	33.9	770	5	AF123884	AF123884 Gallus ga
21	380.5	33.8	1341	5	AF109014	AF109014 Gallus ga
22	378.5	33.6	1268	6	AR023709	AR023709 Sequence
23	378.5	33.6	1268	6	AR225842	AR225842 Sequence
24	378.5	33.6	1675	9	BC008687	BC008687 Homo sapi
25	378.5	33.6	1717	9	BC028226	BC028226 Homo sapi
26	378.5	33.6	79970	2	AC011428	AC011428 Homo sapi
c 27	378.5	33.6	134506	9	AC005738	AC005738 Homo sapi
28	376.5	33.4	1268	9	HSU63842	U63842 Human neuro
29	375.5	33.3	1385	10	MMU76207	U76207 Mus musculu
30	375.5	33.3	10393	10	AF303001	AF303001 Mus muscu
31	374	33.2	258118	2	AC112007	AC112007 Rattus no
c 32	374	33.2	258880	2	AC095430	AC095430 Rattus no
33	370.5	32.9	1412	10	MMMATH4A	Y07621 M.musculus
c 34	370.5	32.9	151044	2	AC011432	AC011432 Homo sapi
35	370	32.8	1527	6	AR308547	AR308547 Sequence
36	370	32.8	1527	10	RNU67777	U67777 Rattus norv
37	368.5	32.7	123855	2	AC102600	AC102600 Mus muscu
38	365	32.4	932	10	MMMATH4C	Y09166 M.musculus
39	365	32.4	1315	10	MMU63841	U63841 Mus musculu
40	365	32.4	1333	6	AR023715	AR023715 Sequence
41	365	32.4	1333	6	AR225848	AR225848 Sequence
42	365	32.4	149268	2	AC124395	AC124395 Mus muscu
43	362	32.1	262798	2	AC111702	AC111702 Rattus no
44	359.5	31.9	735	10	MMU67776	U67776 Mus musculu
45	359.5	31.9	738	6	AR308548	AR308548 Sequence

ALIGNMENTS

RESULT 1

```
HSA133776
LOCUS       HSA133776              1330 bp    DNA     linear   PRI 19-JUN-1999
DEFINITION  Homo sapiens gene for neurogenin 3.
ACCESSION   AJ133776
VERSION     AJ133776.1  GI:5123782
KEYWORDS    bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ravassard,P., Icard-Liepkalns,C., Wiard,L., Julien,J.P. and
            Mallet,J.
  TITLE     The human neurogenin 3 homolog maps to chromosome 10q21.3 and its
            expression pattern is identical to that of its murine counterparts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1330)
  AUTHORS   Ravassard,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-1999) Ravassard P., Lgn, CNRS UMRC 9923, Hopital
            de la Pitie Salpetriere, Bat. CERVI, 83 Bd. de l'Hopital, 75013
            PARIS, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1330
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            1. .1330
                     /gene="ngn3"
     mRNA            join(<1. .157,321. .>1330)
                     /gene="ngn3"
     exon            1. .157
                     /gene="ngn3"
                     /number=1
     intron          158. .320
                     /gene="ngn3"
                     /number=1
     exon            321. .1330
                     /gene="ngn3"
                     /number=2
     CDS             322. .966
                     /gene="ngn3"
                     /function="early neurogenesis"
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="CAB45384.1"
                     /db_xref="GI:5123783"
                     /translation="MTPQPSGAPTVQVTRETERSFPRASEDEVTCPTSAPPSPTRTPG
                     NCAEAEEGGCRGAPRKLRARRGGRSRPKSELALSKQRRSRRKKANDRERNRMHDLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSLYALEPPAPHCGELG
                     SPGGPPGDWGSLYSPVSQAGSLSPAASLEERPGLLGATSSACLSPGSLAFSDFL"
BASE COUNT      230 a    459 c    413 g    228 t
ORIGIN

Alignment Scores:
Pred. No.:              8.59e-47        Length:         1330
Score:                  1127.00         Matches:        214
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-595-947E-10 (1-214) x HSA133776 (1-1330)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 ATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGAGCGGTCC 381

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 TTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCCAGCCCCACT 441

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        442 CGCACACCGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCCCCGAGGAAG 501

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        502 CTCCGGGCACGGCGCGGGGGACGCAGCCGGCCTAAGAGCGAGTTGGCACTGAGCAAGCAG 561

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        562 CGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACGACCTCAAC 621

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        622 TCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCGAAGCTCACC 681

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        682 AAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAACGCTGCGC 741

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        742 ATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGAGCTGGGC 801

Qy        161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        802 AGCCCAGGCGGTCCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCAGGCTGGC 861

Qy        181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        862 AGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCACCTCTTCC 921

Qy        201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
              ||||||||||||||||||||||||||||||||||||||||||
Db        922 GCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTG 963

RESULT 2
AF234829
LOCUS       AF234829               5340 bp    DNA     linear   PRI 19-OCT-2001
DEFINITION  Homo sapiens neurogenin 3 gene, complete cds.
ACCESSION   AF234829
VERSION     AF234829.1  GI:13183002
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5340)
  AUTHORS   del Bosque-Plata,L., Lin,J., Horikawa,Y., Schwarz,P.E., Cox,N.J.,
            Iwasaki,N., Ogata,M., Iwamoto,Y., German,M.S. and Bell,G.I.
  TITLE     Mutations in the coding region of the neurogenin 3 gene (NEUROG3)
            are not a common cause of maturity-onset diabetes of the young in
            Japanese subjects
  JOURNAL   Diabetes 50 (3), 694-696 (2001)
  MEDLINE   21140923
   PUBMED   11246894
REFERENCE   2  (bases 1 to 5340)
  AUTHORS   Lin,J. and German,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2000) Hormone Research Institute, University of
            California San Francisco; 513 Parnassus Ave., San Francisco, CA
            94143-0534, USA
FEATURES             Location/Qualifiers
     source          1. .5340
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     mRNA            <3022. .>3666
                     /product="neurogenin 3"
     CDS             3022. .3666
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="AAK15022.1"
                     /db_xref="GI:13183003"
```

```
                     /translation="MTPQPSGAPTVQVTRETERSFPRASEDEVTCPTSAPPSPTRTRG
                     NCAEAEEGGCRGAPRKLRARRGGRSRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSLYALEPPAPHCGELG
                     SPGGSPGDWGSLYSPVSQAGSLSPAASLEERPGLLGATSSACLSPGSLAFSDFL"
BASE COUNT     1215 a    1500 c    1514 g    1111 t
ORIGIN

Alignment Scores:
Pred. No.:              3.45e-45        Length:         5340
Score:                  1105.00         Matches:        211
Percent Similarity:     99.07%          Conservative:   1
Best Local Similarity:  98.60%          Mismatches:     2
Query Match:            98.05%          Indels:         0
DB:                     9               Gaps:           0

US-09-595-947E-10 (1-214) x AF234829 (1-5340)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3022 ATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGAGCGGTCC 3081

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3082 TTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCCAGCCCCACT 3141

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3142 CGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCCCCGAGGAAG 3201

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3202 CTCCGGGCACGGCGCGGGGGACGCAGCCGGCCTAAGAGCGAGTTGGCACTGAGCAAGCAG 3261

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db       3262 CGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACAACCTCAAC 3321

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3322 TCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCGAAGCTCACC 3381

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3382 AAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAACGCTGCGC 3441

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3442 ATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGAGCTGGGC 3501

Qy        161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db       3502 AGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCAGGCTGGC 3561

Qy        181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3562 AGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCACCTCTTCC 3621

Qy        201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
              ||||||||||||||||||||||||||||||||||||||||||
Db       3622 GCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTG 3663

RESULT 3
AL450311/c
LOCUS       AL450311               165110 bp    DNA     linear   PRI 12-JUL-2001
DEFINITION  Human DNA sequence from clone RP11-343J3 on chromosome 10, complete
            sequence.
ACCESSION   AL450311
VERSION     AL450311.11  GI:14626972
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165110)
  AUTHORS   Howden,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2001 this sequence version replaced gi:14575291.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-343J3 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-343J3 The true
            left end of clone RP11-242G20 is at 139955 in this sequence. The
            true right end of clone RP11-404C6 is at 6588 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .165110
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-343J3"
                     /clone_lib="RPCI-11.2"
     repeat_region   7. .147
                     /note="THE1C repeat: matches 2. .142 of consensus"
     repeat_region   151. .293
                     /note="MIR repeat: matches 2. .148 of consensus"
     repeat_region   1136. .1187
                     /note="26 copies 2 mer gt 98% conserved"
     repeat_region   2960. .3272
                     /note="AluSg repeat: matches 1. .310 of consensus"
     repeat_region   4144. .4270
                     /note="AluSx repeat: matches 1. .134 of consensus"
     repeat_region   4319. .4494
                     /note="AluSx repeat: matches 118. .293 of consensus"
     repeat_region   5662. .6438
                     /note="L1MEc repeat: matches 272. .1095 of consensus"
     repeat_region   6916. .8080
                     /note="L1MEc repeat: matches 1168. .2367 of consensus"
     repeat_region   9934. .10087
                     /note="MIR repeat: matches 25. .184 of consensus"
     repeat_region   10353. .10481
                     /note="MIR repeat: matches 65. .194 of consensus"
     repeat_region   11077. .11372
                     /note="AluSx repeat: matches 1. .294 of consensus"
     repeat_region   11600. .11639
                     /note="20 copies 2 mer tc 95% conserved"
     repeat_region   12479. .12608
                     /note="65 copies 2 mer at 90% conserved"
     repeat_region   13433. .13876
                     /note="MLT1C repeat: matches 1. .466 of consensus"
     repeat_region   14372. .14576
```

```
                     /note="AluSq repeat: matches 1. .301 of consensus"
     repeat_region   14577. .14633
                     /note="AluSq repeat: matches 243. .299 of consensus"
     repeat_region   14843. .14943
                     /note="MER86 repeat: matches 9. .109 of consensus"
     repeat_region   15404. .15613
                     /note="MIR repeat: matches 3. .213 of consensus"
     repeat_region   16626. .16705
                     /note="MIR repeat: matches 60. .140 of consensus"
     repeat_region   16928. .16993
                     /note="33 copies 2 mer gg 66% conserved"
     repeat_region   17476. .17569
                     /note="MER81 repeat: matches 2. .114 of consensus"
     repeat_region   17719. .18069
                     /note="L1MC4 repeat: matches 7617. .7977 of consensus"
     repeat_region   18117. .18345
                     /note="MIR repeat: matches 7. .262 of consensus"
     repeat_region   19169. .19400
                     /note="MIR repeat: matches 5. .239 of consensus"
     repeat_region   19436. .19627
                     /note="L1MC5 repeat: matches 7728. .7917 of consensus"
     repeat_region   19642. .19716
                     /note="MIR repeat: matches 48. .131 of consensus"
     repeat_region   19969. .20010
                     /note="21 copies 2 mer tg 100% conserved"
     repeat_region   20623. .20704
                     /note="41 copies 2 mer gt 85% conserved"
     repeat_region   20738. .20773
                     /note="U2 repeat: matches 1. .36 of consensus"
     repeat_region   20802. .20863
                     /note="MLT1J repeat: matches 1. .62 of consensus"
     repeat_region   20889. .21263
                     /note="MLT1F repeat: matches 188. .541 of consensus"
     repeat_region   21463. .21618
                     /note="MIR repeat: matches 46. .192 of consensus"
     repeat_region   22019. .22326
                     /note="AluJb repeat: matches 1. .306 of consensus"
     repeat_region   22381. .22564
                     /note="L2 repeat: matches 2453. .2629 of consensus"
     repeat_region   22896. .23174
                     /note="MLT1J repeat: matches 117. .413 of consensus"
     repeat_region   23215. .23346
                     /note="MIR repeat: matches 48. .188 of consensus"
     repeat_region   23388. .23532
                     /note="L2 repeat: matches 2097. .2230 of consensus"
     repeat_region   23533. .23837
                     /note="AluJb repeat: matches 1. .303 of consensus"
     repeat_region   23838. .24137
                     /note="L2 repeat: matches 1754. .2097 of consensus"
     repeat_region   24291. .24581
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   24653. .24850
                     /note="MIR repeat: matches 1. .200 of consensus"
     misc_feature    26189. .26685
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    29453. .30918
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   31084. .31156
                     /note="MIR repeat: matches 65. .138 of consensus"
     repeat_region   31618. .31876
                     /note="AluJo repeat: matches 29. .275 of consensus"
     repeat_region   32767. .32830
                     /note="MIR repeat: matches 76. .139 of consensus"
     repeat_region   33050. .33178
                     /note="43 copies 3 mer tcc 72% conserved"
     misc_feature    35112. .36201
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   36551. .36604
                     /note="27 copies 2 mer ac 94% conserved"
     repeat_region   37585. .38254
                     /note="L1PA15 repeat: matches 5480. .6157 of consensus"
     repeat_region   38265. .38561
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   39214. .39423
                     /note="MIR repeat: matches 3. .219 of consensus"
     repeat_region   39489. .39801
                     /note="AluSx repeat: matches 1. .292 of consensus"
     repeat_region   40050. .40189
                     /note="MIR repeat: matches 1. .144 of consensus"
     repeat_region   41057. .41290
                     /note="MIR repeat: matches 8. .240 of consensus"
     repeat_region   41296. .41460
                     /note="L2 repeat: matches 2569. .2730 of consensus"
     repeat_region   41506. .41944
                     /note="L2 repeat: matches 1916. .2416 of consensus"
     repeat_region   42388. .42698
                     /note="L2 repeat: matches 1448. .1779 of consensus"
     repeat_region   44193. .44579
                     /note="THE1C repeat: matches 1. .371 of consensus"
     repeat_region   44600. .44956
                     /note="LTR16A repeat: matches 90. .445 of consensus"
     repeat_region   45240. .45300
                     /note="MER58A repeat: matches 37. .97 of consensus"
     repeat_region   45798. .45909
                     /note="LTR41 repeat: matches 90. .192 of consensus"
     repeat_region   46826. .46871
                     /note="23 copies 2 mer gt 97% conserved"
     repeat_region   50684. .50969
                     /note="AluSq repeat: matches 1. .287 of consensus"
     repeat_region   50980. .51291
                     /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   52222. .52519
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   54065. .54260
                     /note="L1M4 repeat: matches 3865. .4055 of consensus"
     repeat_region   54261. .54432
                     /note="FAM repeat: matches 2. .167 of consensus"
     repeat_region   54433. .54629
                     /note="L1M4 repeat: matches 3652. .3865 of consensus"
     repeat_region   54648. .54862
                     /note="LTR41 repeat: matches 11. .217 of consensus"
     repeat_region   54863. .55236
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   55237. .55700

Alignment Scores:
Pred. No.:              1.42e-43        Length:         165110
Score:                  1099.00         Matches:        210
Percent Similarity:     98.60%          Conservative:   1
Best Local Similarity:  98.13%          Mismatches:     3
Query Match:            97.52%          Indels:         0
DB:                     9               Gaps:           0

US-09-595-947E-10 (1-214) x AL450311 (1-165110)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30325 ATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGAGCGGTCC 30266

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30265 TTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCCAGCCCCACT 30206

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30205 CGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCCCCGAGGAAG 30146

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30145 CTCCGGGCACGGCGCGGGGGACGCAGCCGGCCTAAGAGCGAGTTGGCACTGAGCAAGCAG 30086

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
```

```
Db      30085 CGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACAACCTCAAC 30026

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30025 TCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCGAAGCTCACC 29966

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29965 AAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAACGCTGCGC 29906

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      29905 ATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGAGCTGGGC 29846

Qy        161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
              ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db      29845 AGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCAGGCTGGC 29786

Qy        181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db      29785 AGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCACCTTTTCC 29726

Qy        201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
              |||||||||||||||||||||||||||||||||||||||||||
Db      29725 GCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTG 29684
```

```
RESULT 4
AC021954/c
LOCUS       AC021954               173341 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-57E12 map 10, WORKING DRAFT
            SEQUENCE, 24 unordered pieces.
ACCESSION   AC021954
VERSION     AC021954.3  GI:7417809
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173341)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 10, clone RP11-57E12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173341)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 173341)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 5, 2000 this sequence version replaced gi:6984451.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
               Center project name: L5931
               Center clone name: 57_E_12
            -------------- Summary Statistics
               Sequencing vector: M13; M77815; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 161190 bases at least Q40
               Consensus quality: 166837 bases at least Q30
               Consensus quality: 168995 bases at least Q20
               Insert size: 176000; agarose-fp
               Insert size: 171041; sum-of-contigs
               Quality coverage: 3.7 in Q20 bases; agarose-fp
               Quality coverage: 3.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 24 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1574: contig of 1574 bp in length
            *     1575     1674: gap of 100 bp
            *     1675     3043: contig of 1369 bp in length
            *     3044     3143: gap of 100 bp
            *     3144     5889: contig of 2746 bp in length
            *     5890     5989: gap of 100 bp
            *     5990     8979: contig of 2990 bp in length
            *     8980     9079: gap of 100 bp
            *     9080    13674: contig of 4595 bp in length
            *    13675    13774: gap of 100 bp
            *    13775    18831: contig of 5057 bp in length
            *    18832    18931: gap of 100 bp
            *    18932    23526: contig of 4595 bp in length
            *    23527    23626: gap of 100 bp
            *    23627    27386: contig of 3760 bp in length
            *    27387    27486: gap of 100 bp
            *    27487    32572: contig of 5086 bp in length
            *    32573    32672: gap of 100 bp
            *    32673    38632: contig of 5960 bp in length
            *    38633    38732: gap of 100 bp
            *    38733    43735: contig of 5003 bp in length
            *    43736    43835: gap of 100 bp
            *    43836    49020: contig of 5185 bp in length
            *    49021    49120: gap of 100 bp
            *    49121    53660: contig of 4540 bp in length
            *    53661    53760: gap of 100 bp
```

```
            *    53761    59544: contig of 5784 bp in length
            *    59545    59644: gap of 100 bp
            *    59645    66395: contig of 6751 bp in length
            *    66396    66495: gap of 100 bp
            *    66496    74696: contig of 8201 bp in length
            *    74697    74796: gap of 100 bp
            *    74797    81326: contig of 6530 bp in length
            *    81327    81426: gap of 100 bp
            *    81427    89446: contig of 8020 bp in length
            *    89447    89546: gap of 100 bp
            *    89547    96459: contig of 6913 bp in length
            *    96460    96559: gap of 100 bp
            *    96560   106498: contig of 9939 bp in length
            *   106499   106598: gap of 100 bp
            *   106599   116986: contig of 10388 bp in length
            *   116987   117086: gap of 100 bp
            *   117087   128890: contig of 11804 bp in length
            *   128891   128990: gap of 100 bp
            *   128991   147290: contig of 18300 bp in length
            *   147291   147390: gap of 100 bp
            *   147391   173341: contig of 25951 bp in length.
FEATURES             Location/Qualifiers
     source          1. .173341
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10"
                     /clone="RP11-57E12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1574
                     /note="assembly_fragment"
     misc_feature    1675. .3043
                     /note="assembly_fragment"
     misc_feature    3144. .5889
                     /note="assembly_fragment"
     misc_feature    5990. .8979
                     /note="assembly_fragment"
     misc_feature    9080. .13674
                     /note="assembly_fragment"
     misc_feature    13775. .18831
                     /note="assembly_fragment"
     misc_feature    18932. .23526
                     /note="assembly_fragment"
     misc_feature    23627. .27386
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    27487. .32572
                     /note="assembly_fragment"
     misc_feature    32673. .38632
                     /note="assembly_fragment"
     misc_feature    38733. .43735
                     /note="assembly_fragment"
     misc_feature    43836. .49020
                     /note="assembly_fragment"
     misc_feature    49121. .53660
                     /note="assembly_fragment"
     misc_feature    53761. .59544
                     /note="assembly_fragment"
     misc_feature    59645. .66395
                     /note="assembly_fragment"
     misc_feature    66496. .74696
                     /note="assembly_fragment"
     misc_feature    74797. .81326
                     /note="assembly_fragment"
     misc_feature    81427. .89446
                     /note="assembly_fragment"
     misc_feature    89547. .96459
                     /note="assembly_fragment"
     misc_feature    96560. .106498
                     /note="assembly_fragment"
     misc_feature    106599. .116986
                     /note="assembly_fragment"
     misc_feature    117087. .128890
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    128991. .147290
                     /note="assembly_fragment"
     misc_feature    147391. .173341
                     /note="assembly_fragment"

Alignment Scores:
Pred. No.:              4.54e-43        Length:         173341
Score:                  1089.00         Matches:        209
Percent Similarity:     98.13%          Conservative:   1
Best Local Similarity:  97.66%          Mismatches:     4
Query Match:            96.63%          Indels:         0
DB:                     2               Gaps:           0

US-09-595-947E-10 (1-214) x AC021954 (1-173341)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6765 ATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGAGCGGTCC 6706

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6705 TTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCCAGCCCCACT 6646

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6645 CGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCCCCGAGGAAG 6586

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||||||
Db       6585 CTCCGGGCACGGCGCGGGGGACGCAGCCGGCTTAAGAGCGAGTTGGCACTGAGCAAGCAG 6526

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       6525 CGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACAACCTCAAC 6466

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6465 TCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCGAAGCTCACC 6406

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6405 AAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAACGCTGCGC 6346

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6345 ATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGAGCTGGGC 6286

Qy        161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
              ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6285 AGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCAGGCTGGC 6226

Qy        181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db       6225 AGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCACCTTTTCC 6166

Qy        201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
              ||||||||||||||||||||||||||||||||||||||||||
Db       6165 GCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTG 6124

RESULT 5
A91167
LOCUS       A91167                 1491 bp    DNA     linear   PAT 22-JAN-2000
DEFINITION  Sequence 1 from Patent WO9827206.
ACCESSION   A91167
VERSION     A91167.1  GI:6740202
KEYWORDS    .
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1491)
  AUTHORS   Icard-Liepkalns,C., Mallet,J. and Corresponding,N.A.
  JOURNAL   Patent: WO 9827206-A 1 25-JUN-1998;
            ICARD LIEPKALNS CHRISTINE (FR); MALLET JACQUES (FR)
FEATURES             Location/Qualifiers
     source          1. .1491
                     /organism="Rattus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10118"
     CDS             459. .1103
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAB69414.1"
                     /db_xref="GI:6740203"
                     /translation="MAPHPLDAPTIQVSQETQQPFPGASDHEVLSSNSTPPSPTLVPR
                     DCSEAEAGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSSGDWGSIYSPVSQAGSLSPTASLEEFPGLQVPSSPSCLLPGTLVFSDFL"
BASE COUNT      307 a    487 c    413 g    284 t
ORIGIN

Alignment Scores:
Pred. No.:                3.02e-33        Length:          1491
Score:                    849.00          Matches:         166
Percent Similarity:       83.72%          Conservative:    14
Best Local Similarity:    77.21%          Mismatches:      33
Query Match:              75.33%          Indels:          2
DB:                       6               Gaps:            2

US-09-595-947E-10 (1-214) x A91167 (1-1491)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||||||||:::||||||:::::::||||||::::::
Db        459 ATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCC 518

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||  ||||||:::  ||||||           |||   ||||||||||||||||
Db        519 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 578

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                 |||   :::|||:::||||||||||   |||   |||||||||      ||||||
Db        579 CTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAG 638

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        639 CTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAG 698

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        699 CGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAAC 758

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        759 TCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 818

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        819 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGC 878

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||      ||||||||   |||   ||||||||||||||||
Db        879 ATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC---TGTGGGGAGCTGGGA 935

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||  ||||||       |||||||||||:::||||||||||||||||||||||||
Db        936 AGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCT 995

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              ||||||||||||||   ||||||||||||||   |||||||||      :::|||
Db        996 GGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCC 1055

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
              :::||||||   ||||||:::|||   ||||||||||||||||
Db       1056 CCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTG 1100

RESULT 6
BD023626
LOCUS       BD023626                1491 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Polypeptide belonging to the family of basic helix-loop-helix
            (bHLH) family and nucleic acid sequence corresponding thereto.
ACCESSION   BD023626
VERSION     BD023626.1  GI:22564849
KEYWORDS    JP 2001510464-A/1.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1491)
  AUTHORS   Liepkalns,C.I., Mallet,J. and Ravassard,P.
  TITLE     Polypeptide belonging to the family of basic helix-loop-helix
            (bHLH) family and nucleic acid sequence corresponding thereto
  JOURNAL   Patent: JP 2001510464-A 1 31-JUL-2001;
            RHONE POULENC RORER SA
COMMENT     OS   Rattus sp. (rat)
            PN   JP 2001510464-A/1
            PD   31-JUL-2001
            PF   19-DEC-1997 JP 1998527415
            PR   19-DEC-1996 FR     96/15651
            PI   CHRISTINE ICARD LIEPKALNS,JACQUES MALLET,PHILIPPE RAVASSARD PC
              C07K14/47,A61K31/711,A61K35/76,A61K38/00,A61K48/00,A61P25/00, PC
             C12N15/09,
            PC   C12N15/00,A61K37/02
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES             Location/Qualifiers
     source          1. .1491
                     /organism="Rattus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10118"
BASE COUNT      307 a    487 c    413 g    284 t
ORIGIN

Alignment Scores:
Pred. No.:                3.02e-33        Length:          1491
Score:                    849.00          Matches:         166
Percent Similarity:       83.72%          Conservative:    14
Best Local Similarity:    77.21%          Mismatches:      33
Query Match:              75.33%          Indels:          2
DB:                       6               Gaps:            2

US-09-595-947E-10 (1-214) x BD023626 (1-1491)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||||||||:::||||||:::::::||||||::::::
Db        459 ATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCC 518

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||  ||||||:::  ||||||           |||   ||||||||||||||||
Db        519 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 578

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                 |||   :::|||:::||||||||||   |||   |||||||||      ||||||
Db        579 CTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAG 638

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        639 CTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAG 698

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        699 CGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAAC 758
```

```
Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        759 TCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 818

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        819 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGC 878

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              |||||||||||||||   |||       |||||||||   |||    ||||||||||||||
Db        879 ATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC---TGTGGGGAGCTGGGA 935

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::|||||||||||||||||||||
Db        936 AGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCT 995

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||||||||||||||   ||||||||||||||   |||||||||           :::|||
Db        996 GGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCC 1055

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                 :::||||||   ||||||:::|||   ||||||||||||||
Db       1056 CCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTG 1100

RESULT 7
RNRELAXT
LOCUS       RNRELAXT                1491 bp    mRNA    linear   ROD 06-MAY-1997
DEFINITION  R.norvegicus mRNA for transcriptional regulator, Relax.
ACCESSION   Y10619
VERSION     Y10619.1  GI:2072737
KEYWORDS    Relax; transcriptional regulator.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Ravassard,P., Chatail,F., Mallet,J. and Icard-Liepkalns,C.
  TITLE     Relax, a novel rat bHLH transcriptional regulator transiently
            expressed in the ventricular proliferating zone of the developing
            central nervous system
  JOURNAL   J. Neurosci. Res. 48 (2), 146-158 (1997)
  MEDLINE   97276390
   PUBMED   9130143
REFERENCE   2  (bases 1 to 1491)
  AUTHORS   Ravassard,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,
            Hopital de la Pitie Salpetriere, 83 Bd. de l'Hopital, F-75013
            Paris, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1491
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone_lib="RT-PCR"
     CDS             459. .1103
                     /function="transcriptional regulator"
                     /codon_start=1
                     /product="Relax"
                     /protein_id="CAA71630.1"
                     /db_xref="GI:2072738"
                     /db_xref="SPTREMBL:O08718"
                     /translation="MAPHPLDAPTIQVSQETQQPFPGASDHEVLSSNSTPPSPTLVPR
                     DCSEAEAGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSSGDWGSIYSPVSQAGSLSPTASLEEFPGLQVPSSPSCLLPGTLVFSDFL"
BASE COUNT      307 a    487 c    413 g    284 t
ORIGIN

Alignment Scores:

Pred. No.:              3.02e-33        Length:         1491
Score:                  849.00          Matches:        166
Percent Similarity:     83.72%          Conservative:   14
Best Local Similarity:  77.21%          Mismatches:     33
Query Match:            75.33%          Indels:         2
DB:                     10              Gaps:           2

US-09-595-947E-10 (1-214) x RNRELAXT (1-1491)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||||||||:::||||||::::::||||||::::::
Db        459 ATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCC 518

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   |||||||:::   ||||||              |||   ||||||||||||||
Db        519 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 578

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                    |||   :::|||:::|||||||||   |||   |||||||||       ||||||
Db        579 CTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAG 638

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        639 CTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAG 698

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        699 CGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAAC 758

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        759 TCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 818

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        819 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGC 878

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              |||||||||||||||   |||       |||||||||   |||    ||||||||||||||
Db        879 ATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC---TGTGGGGAGCTGGGA 935

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::|||||||||||||||||||||
Db        936 AGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCT 995

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||||||||||||||   |||||||||||||||   |||||||||           :::|||
Db        996 GGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCC 1055

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                 :::||||||   ||||||:::|||   ||||||||||||||
Db       1056 CCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTG 1100

RESULT 8
AC127817
LOCUS       AC127817                258815 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-259G16, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC127817
VERSION     AC127817.3  GI:25077905
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 258815)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
```

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Pan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 258815)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 258815)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 19, 2002 this sequence version replaced gi:23912578.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GZXS
Center clone name: CH230-259G16
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 224747 bases at least Q40
Consensus quality: 227981 bases at least Q30
Consensus quality: 229752 bases at least Q20
Estimated insert size: 228243; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 255979: contig of 255979 bp in length
* 255980 256079: gap of unknown length
* 256080 257349: contig of 1270 bp in length
* 257350 257449: gap of unknown length
* 257450 258815: contig of 1366 bp in length.

FEATURES Location/Qualifiers
source 1. .258815
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-259G16"
misc_feature 1. .1609
/note="wgs_end_extension
clone_end:Sp6"
misc_feature complement(2865. .3659)
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BZ169379"
misc_feature 224098. .225157
/note="clone_boundary
clone_end:T7
site:
end_sequence:BZ169378"
BASE COUNT 59351 a 56888 c 58131 g 56821 t 27624 others
ORIGIN

Alignment Scores:

Pred. No.:	2.94e-31	Length:	258815
Score:	849.00	Matches:	166
Percent Similarity:	83.72%	Conservative:	14
Best Local Similarity:	77.21%	Mismatches:	33
Query Match:	75.33%	Indels:	2
DB:	2	Gaps:	2

US-09-595-947E-10 (1-214) x AC127817 (1-258815)

```
Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||      |||||||||:::||||||::::::||||||::::::
Db     247860 ATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCC 247919

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||  ||||||:::  ||||||          |||   ||||||||||||||||
Db     247920 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 247979

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                 |||   :::|||:::|||||||||   |||   |||||||||      ||||||
Db     247980 CTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAG 248039

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
```

```
                ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db      248040 CTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAG 248099

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db      248100 CGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAAC 248159

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248160 TCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 248219

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248220 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGC 248279

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||       |||||||||   |||   ||||||||||||||||
Db      248280 ATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC---TGTGGGGAGCTGGGA 248336

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::||||||||||||||||||||
Db      248337 AGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCT 248396

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||||||||||||||   ||||||||||||||   |||||||||              :::|||
Db      248397 GGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCC 248456

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                 :::||||||   ||||||:::|||   ||||||||||||||
Db      248457 CCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTG 248501

RESULT 9
AX698801
LOCUS       AX698801                861 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 7 from Patent WO02086107.
ACCESSION   AX698801
VERSION     AX698801.1  GI:29499589
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Wobus,A.M., St-Onge,L., Blyszczuk,P. and Hoffmann,U.
  TITLE     A method for differentiating stem cells into insulin-producing
            cells
  JOURNAL   Patent: WO 02086107-A 7 31-OCT-2002;
            DeveloGen Aktiengesellschaft fuer Entwicklungsbiologische Forschung
            (DE) ; INSTITUT FUER PFLANZENGENETIK UND KULTURPFLANZENFORSCHUNG
            (DE)
FEATURES             Location/Qualifiers
     source          1. .861
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     CDS             160. .804
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD87442.1"
                     /db_xref="GI:29499590"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT      182 a    274 c    250 g    155 t
ORIGIN

Alignment Scores:
Pred. No.:              5.32e-32        Length:         861
Score:                  819.00          Matches:        163
Percent Similarity:     82.33%          Conservative:   14
Best Local Similarity:  75.81%          Mismatches:     36
Query Match:            72.67%          Indels:         2

DB:                     6               Gaps:           2

US-09-595-947E-10 (1-214) x AX698801 (1-861)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::||||||:::   ||||||::::::
Db        160 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 219

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||              |||   ||||||||||||||
Db        220 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 279

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                 |||   :::|||:::|||||||||   |||   |||||||||       ||||||
Db        280 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 339

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        340 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 399

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        400 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 459

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 519

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 579

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              |||||||||||||||   |||       |||||||||   |||   |||||||||||||||
Db        580 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 636

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::||||||||||||||||||||
Db        637 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 696

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   |||||||||||||||   |||||||||              :::|||
Db        697 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 756

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                 :::   |||   ||||||:::|||   ||||||||||||||
Db        757 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 801

RESULT 10
MMU76208
LOCUS       MMU76208                861 bp    DNA     linear   ROD 05-FEB-1997
DEFINITION  Mus musculus neurogenin 3 (ngn3) gene, complete cds.
ACCESSION   U76208
VERSION     U76208.1  GI:1815654
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 861)
  AUTHORS   Sommer,L., Ma,Q. and Anderson,D.J.
  TITLE     neurogenins, a novel family of atonal-related bHLH transcription
            factors, are putative mammalian neuronal determination genes that
            reveal progenitor cell heterogeneity in the developing CNS and PNS
  JOURNAL   Mol. Cell. Neurosci. 8 (4), 221-241 (1996)
  MEDLINE   97153565
   PUBMED   9000438
REFERENCE   2  (bases 1 to 861)
  AUTHORS   Sommer,L., Ma,Q. and Anderson,D.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-1996) Biology 216-76, California Institute of
            Technology, Howard Hughes Medical Institute, Pasadena, CA 91125,
            USA
```

```
COMMENT     On Feb 5, 1997 this sequence version replaced gi:1666911.
FEATURES             Location/Qualifiers
     source          1. .861
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene            160. .804
                     /gene="ngn3"
     CDS             160. .804
                     /gene="ngn3"
                     /function="bHLH transcription factor"
                     /note="NGN3"
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="AAC53029.1"
                     /db_xref="GI:1666912"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT      182 a    274 c    250 g    155 t
ORIGIN

Alignment Scores:
Pred. No.:              5.32e-32      Length:         861
Score:                  819.00        Matches:        163
Percent Similarity:     82.33%        Conservative:   14
Best Local Similarity:  75.81%        Mismatches:     36
Query Match:            72.67%        Indels:         2
DB:                     10            Gaps:           2

US-09-595-947E-10 (1-214) x MMU76208 (1-861)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::||||||:::   ||||||::::::
Db        160 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 219

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||  ||||||:::  ||||||          |||   ||||||||||||||||
Db        220 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 279

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                  |||   :::|||:::|||||||||   |||   |||||||||      ||||||
Db        280 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 339

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        340 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 399

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        400 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 459

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 519

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 579

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||       |||||||||   |||   ||||||||||||||
Db        580 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 636

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::||||||||||||||||||||
Db        637 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 696

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   ||||||||||||||   |||||||||          :::|||
Db        697 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 756

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                  :::   |||   ||||||:::|||   ||||||||||||||
Db        757 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 801

RESULT 11
AF364300
LOCUS       AF364300               1861 bp    DNA     linear   ROD 03-MAY-2001
DEFINITION  Mus musculus neurogenin 3 gene, complete cds.
ACCESSION   AF364300
VERSION     AF364300.1  GI:13937128
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1861)
  AUTHORS   Lee,J., Smith,S., Watada,H., Lin,J., Scheel,D., Wang,J., Mirmira,R.
            and German,M.
  TITLE     Regulation of the pancreatic pro-endocrine gene neurogenin3
  JOURNAL   Diabetes (2001) In press
REFERENCE   2  (bases 1 to 1861)
  AUTHORS   Schwitzgebel,V. and German,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2001) Hormone Research Institute, University of
            California San Francisco, 513 Parnassus Ave, HSW1090, San
            Francisco, CA 94145-0534, USA
FEATURES             Location/Qualifiers
     source          1. .1861
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     mRNA            <1093. .>1737
                     /product="neurogenin 3"
     CDS             1093. .1737
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="AAK50058.1"
                     /db_xref="GI:13937129"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT      397 a    560 c    537 g    367 t
ORIGIN

Alignment Scores:
Pred. No.:              1.05e-31      Length:         1861
Score:                  819.00        Matches:        163
Percent Similarity:     82.33%        Conservative:   14
Best Local Similarity:  75.81%        Mismatches:     36
Query Match:            72.67%        Indels:         2
DB:                     10            Gaps:           2

US-09-595-947E-10 (1-214) x AF364300 (1-1861)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::||||||:::   ||||||::::::
Db       1093 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 1152

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||  ||||||:::  ||||||          |||   ||||||||||||||||
Db       1153 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 1212

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                  |||   :::|||:::|||||||||   |||   |||||||||      ||||||
Db       1213 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 1272

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db       1273 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 1332

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
```

```
Db       1333 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 1392

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1393 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 1452

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1453 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 1512

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||      |||||||||   |||   ||||||||||||||||
Db       1513 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 1569

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::||||||||||||||||||||
Db       1570 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 1629

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   ||||||||||||||   |||||||||         :::|||
Db       1630 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 1689

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                 :::   |||   ||||||:::|||   ||||||||||||||
Db       1690 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 1734

RESULT 12
MMMATH4B
LOCUS       MMMATH4B                5567 bp    DNA     linear   ROD 25-OCT-2000
DEFINITION  M.musculus MATH4B gene.
ACCESSION   Y09167
VERSION     Y09167.2  GI:11065737
KEYWORDS    bHLH; mammalian atonal homolog 4B; MATH4B; neurogenin 3.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Cau,E., Gradwohl,G., Fode,C. and Guillemot,F.
  TITLE     Mash1 activates a cascade of bHLH regulators in olfactory neuron
            progenitors
  JOURNAL   Development 124 (8), 1611-1621 (1997)
  MEDLINE   97261963
   PUBMED   9108377
REFERENCE   2
  AUTHORS   Jacquemin,P., Durviaux,S.M., Jensen,J., Godfraind,C., Gradwohl,G.,
            Guillemot,F., Madsen,O.D., Carmeliet,P., Dewerchin,M., Collen,D.,
            Rousseau,G.G. and Lemaigre,F.P.
  TITLE     Transcription factor hepatocyte nuclear factor 6 regulates
            pancreatic endocrine cell differentiation and controls expression
            of the proendocrine gene ngn3
  JOURNAL   Mol. Cell. Biol. 20 (12), 4445-4454 (2000)
  MEDLINE   20285449
   PUBMED   10825208
REFERENCE   3
  AUTHORS   Gradwohl,G.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC,
            CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg,
            F-67404 ILLKIRCH cedex, FRANCE
  REMARK    Revised by [4]
REFERENCE   4  (bases 1 to 5567)
  AUTHORS   Lemaigre,F.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic
            Research Unit, Louvain University Medical School, Avenue Hippocrate
            75, box 7529, Brussels 1200, BELGIUM
COMMENT     On Oct 31, 2000 this sequence version replaced gi:1666087.
FEATURES             Location/Qualifiers
     source          1. .5567
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="SV129"
                     /db_xref="taxon:10090"
                     /clone="lambda GEM12 clone 19C"
                     /cell_line="D3"
                     /cell_type="embryonic stem cells"
                     /clone_lib="Genomic library No.62, MboI partial digests
                     cloned in BamHI"
                     /dev_stage="embryos"
     misc_binding    958
                     /bound_moiety="HNF6"
     misc_binding    3692
                     /bound_moiety="HNF6"
     gene            4145. .5567
                     /gene="MATH4B"
     TATA_signal     4145. .4151
                     /gene="MATH4B"
                     /note="putative"
     CDS             4923. .5567
                     /gene="MATH4B"
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="CAA70366.1"
                     /db_xref="GI:1666088"
                     /db_xref="SWISS-PROT:P70661"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT     1271 a   1549 c   1564 g   1183 t
ORIGIN

Alignment Scores:
Pred. No.:              2.79e-31        Length:         5567
Score:                  819.00          Matches:        163
Percent Similarity:     82.33%          Conservative:   14
Best Local Similarity:  75.81%          Mismatches:     36
Query Match:            72.67%          Indels:         2
DB:                     10              Gaps:           2

US-09-595-947E-10 (1-214) x MMMATH4B (1-5567)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::||||||:::   ||||||::::::
Db       4923 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 4982

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||              |||   ||||||||||||||||
Db       4983 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 5042

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                 |||   :::|||:::|||||||||   |||   |||||||||       ||||||
Db       5043 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 5102

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db       5103 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 5162

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       5163 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 5222

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5223 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 5282

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5283 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 5342

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||      |||||||||   |||   ||||||||||||||||
Db       5343 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 5399

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
```

```
                ||||||   ||||||      ||||||||||||||:::||||||||||||||||||||
Db       5400 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 5459

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuGlyAlaThrSer 199
              |||:::|||||||||   ||||||||||||||   |||||||||          :::|||
Db       5460 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 5519

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                  :::   |||   ||||||:::|||   ||||||||||||||
Db       5520 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 5564

RESULT 13
AC109783/c
LOCUS       AC109783               138070 bp    DNA     linear   HTG 07-FEB-2002
DEFINITION  Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
            unordered pieces.
ACCESSION   AC109783
VERSION     AC109783.1  GI:18581594
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 138070)
  AUTHORS   McCombie,W.R., Baker,J.P., Balija,V., Dedhia,N.N., de la
            Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
            Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,
            Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
            Zutavern,T.
  TITLE     Mouse Genomic Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 138070)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
COMMENT     -------------- Genome Center --------------
            Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
            Laboratory
            Center code: CSHL
            Web site: http://www.cshl.org/genseq
            Contact:  mccombie@cshl.org
            ------------------- Project Information
            Project name: RP23-121F10
            Clone name: RP23-121F10
            Insert size: 173000; agarose-fp
            Insert size: 141616; sum-of-contigs
            Quality coverage: 4.00 in Q20 bases; agarose-fp
            Quality coverage: 3.70 in Q20 bases; sum-of-contigs
            ----------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17785: contig of 17785 bp in length
            *    17786    17874: gap of unknown length
            *    17875    30356: contig of 12482 bp in length
            *    30357    30444: gap of unknown length
            *    30445    42306: contig of 11862 bp in length
            *    42307    42394: gap of unknown length
            *    42395    53598: contig of 11204 bp in length
            *    53599    53686: gap of unknown length
            *    53687    64355: contig of 10669 bp in length
            *    64356    64443: gap of unknown length
            *    64444    74016: contig of 9573 bp in length
            *    74017    74104: gap of unknown length
            *    74105    83366: contig of 9262 bp in length
            *    83367    83454: gap of unknown length
            *    83455    92355: contig of 8901 bp in length
            *    92356    92443: gap of unknown length
            *    92444   100821: contig of 8378 bp in length
            *   100822   100909: gap of unknown length
            *   100910   107529: contig of 6620 bp in length
            *   107530   107617: gap of unknown length
            *   107618   114066: contig of 6449 bp in length
            *   114067   114154: gap of unknown length
            *   114155   118873: contig of 4719 bp in length
            *   118874   118961: gap of unknown length
            *   118962   123619: contig of 4658 bp in length
            *   123620   123707: gap of unknown length
            *   123708   128240: contig of 4533 bp in length
            *   128241   128328: gap of unknown length
            *   128329   132682: contig of 4354 bp in length
            *   132683   132770: gap of unknown length
            *   132771   136341: contig of 3571 bp in length
            *   136342   136429: gap of unknown length
            *   136430   138070: contig of 1641 bp in length.
FEATURES             Location/Qualifiers
     source          1. .138070
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-121F10"
BASE COUNT    35731 a  33657 c  33303 g  33954 t   1425 others
ORIGIN

Alignment Scores:
Pred. No.:              4.82e-30        Length:         138070
Score:                  819.00          Matches:        163
Percent Similarity:     82.33%          Conservative:   14
Best Local Similarity:  75.81%          Mismatches:     36
Query Match:            72.67%          Indels:         2
DB:                     2               Gaps:           2

US-09-595-947E-10 (1-214) x AC109783 (1-138070)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::||||||:::   ||||||::::::
Db     111220 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 111161

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   |||||             |||   ||||||||||||||
Db     111160 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 111101

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                 |||   :::|||:::|||||||||   |||   ||||||||       ||||||
Db     111100 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 111041

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db     111040 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 110981

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db     110980 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 110921

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     110920 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 110861

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     110860 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 110801

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||       |||||||||   |||   ||||||||||||||
Db     110800 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 110744

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::||||||||||||||||||||
Db     110743 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 110684
```

```
Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   |||||||||||||||   |||||||||          :::|||
Db     110683 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 110624

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                  :::   |||   ||||||:::|||   ||||||||||||||
Db     110623 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 110579

RESULT 14
AC127417
LOCUS       AC127417               215050 bp    DNA     linear   HTG 19-OCT-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-459M2, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC127417
VERSION     AC127417.2  GI:24137619
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 215050)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 215050)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 215050)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Oct 19, 2002 this sequence version replaced gi:21759524.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BA0459M02
            ----------
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 212279 bases at least Q40
            Consensus quality: 213137 bases at least Q30
            Consensus quality: 213590 bases at least Q20
            Insert size: 192000; agarose-fp
            Insert size: 217944; sum-of-contigs
            Quality coverage: 11.89 in Q20 bases; agarose-fp
            Quality coverage: 10.54 in Q20 bases; sum-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4227: contig of 4227 bp in length
            *     4228     4327: gap of unknown length
            *     4328    16589: contig of 12262 bp in length
            *    16590    16689: gap of unknown length
            *    16690    27318: contig of 10629 bp in length
            *    27319    27418: gap of unknown length
            *    27419    49797: contig of 22379 bp in length
            *    49798    49897: gap of unknown length
            *    49898    79378: contig of 29481 bp in length
            *    79379    79478: gap of unknown length
            *    79479   132595: contig of 53117 bp in length
            *   132596   132695: gap of unknown length
            *   132696   215050: contig of 82355 bp in length.
FEATURES             Location/Qualifiers
     source          1. .215050
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-459M2"
     misc_feature    1. .4227
                     /note="assembly_name:Contig31"
     misc_feature    4328. .16589
                     /note="assembly_name:Contig32"
     misc_feature    16690. .27318
                     /note="assembly_name:Contig33"
     misc_feature    27419. .49797
                     /note="assembly_name:Contig34"
     misc_feature    49898. .79378
                     /note="assembly_name:Contig35"
     misc_feature    79479. .132595
                     /note="assembly_name:Contig36"
     misc_feature    132696. .215050
                     /note="assembly_name:Contig37"
BASE COUNT    56024 a  52513 c  52885 g  53012 t    616 others
ORIGIN

Alignment Scores:
Pred. No.:                7.14e-30      Length:          215050
Score:                    819.00        Matches:         163
Percent Similarity:       82.33%        Conservative:    14
Best Local Similarity:    75.81%        Mismatches:      36
Query Match:              72.67%        Indels:          2
DB:                       2             Gaps:            2

US-09-595-947E-10 (1-214) x AC127417 (1-215050)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::|||||:::   ||||||::::::
Db     205567 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 205626

Qy         21 ProProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||              |||   ||||||||||||||||
Db     205627 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 205686

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                  |||   :::|||:::|||||||||   |||   |||||||||       ||||||
Db     205687 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 205746

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db     205747 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 205806

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db     205807 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 205866

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205867 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 205926

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205927 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 205986

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||   |||||||||   |||   ||||||||||||||||
Db     205987 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 206043
```

```
Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::|||||||||||||||||||
Db     206044 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 206103

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   |||||||||||||||   |||||||||          :::|||
Db     206104 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 206163

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                  :::   |||   ||||||:::|||   ||||||||||||||
Db     206164 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 206208

RESULT 15
AC011010/c
LOCUS       AC011010               170896 bp    DNA     linear   HTG 16-MAR-2000
DEFINITION  Homo sapiens clone RP11-6P16, WORKING DRAFT SEQUENCE, 21 unordered
            pieces.
ACCESSION   AC011010
VERSION     AC011010.4  GI:7107881
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170896)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-6P16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170896)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2000 this sequence version replaced gi:6479051.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2916
                Center clone name: 6_P_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 114103 bases at least Q40
                Consensus quality: 141555 bases at least Q30
                Consensus quality: 158230 bases at least Q20
                Insert size: 154000; agarose-fp
                Insert size: 168896; sum-of-contigs
                Quality coverage: 3.6 in Q20 bases; agarose-fp
                Quality coverage: 3.3 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1744: contig of 1744 bp in length
            *    1745     1844: gap of 100 bp
            *    1845     2954: contig of 1110 bp in length
            *    2955     3054: gap of 100 bp
            *    3055     4839: contig of 1785 bp in length
            *    4840     4939: gap of 100 bp
            *    4940     7461: contig of 2522 bp in length
            *    7462     7561: gap of 100 bp
            *    7562    11912: contig of 4351 bp in length
            *   11913    12012: gap of 100 bp
            *   12013    15127: contig of 3115 bp in length
            *   15128    15227: gap of 100 bp
            *   15228    19176: contig of 3949 bp in length
            *   19177    19276: gap of 100 bp
            *   19277    24838: contig of 5562 bp in length
            *   24839    24938: gap of 100 bp
            *   24939    30632: contig of 5694 bp in length
            *   30633    30732: gap of 100 bp
            *   30733    37821: contig of 7089 bp in length
            *   37822    37921: gap of 100 bp
            *   37922    44638: contig of 6717 bp in length
            *   44639    44738: gap of 100 bp
            *   44739    49424: contig of 4686 bp in length
            *   49425    49524: gap of 100 bp
            *   49525    56517: contig of 6993 bp in length
            *   56518    56617: gap of 100 bp
            *   56618    65413: contig of 8796 bp in length
            *   65414    65513: gap of 100 bp
            *   65514    75659: contig of 10146 bp in length
            *   75660    75759: gap of 100 bp
            *   75760    86433: contig of 10674 bp in length
            *   86434    86533: gap of 100 bp
            *   86534    98763: contig of 12230 bp in length
            *   98764    98863: gap of 100 bp
            *   98864   109919: contig of 11056 bp in length
            *  109920   110019: gap of 100 bp
            *  110020   125264: contig of 15245 bp in length
            *  125265   125364: gap of 100 bp
            *  125365   141043: contig of 15679 bp in length
            *  141044   141143: gap of 100 bp
            *  141144   170896: contig of 29753 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170896
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-6P16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1744
                     /note="assembly_fragment"
     misc_feature    1845. .2954
                     /note="assembly_fragment"
     misc_feature    3055. .4839
                     /note="assembly_fragment"
     misc_feature    4940. .7461
                     /note="assembly_fragment"
     misc_feature    7562. .11912
                     /note="assembly_fragment"
     misc_feature    12013. .15127
                     /note="assembly_fragment"
     misc_feature    15228. .19176
                     /note="assembly_fragment"
     misc_feature    19277. .24838
                     /note="assembly_fragment"
     misc_feature    24939. .30632
                     /note="assembly_fragment"
     misc_feature    30733. .37821
                     /note="assembly_fragment"
```

```
     misc_feature    37922. .44638
                     /note="assembly_fragment"
     misc_feature    44739. .49424
                     /note="assembly_fragment"
     misc_feature    49525. .56517
                     /note="assembly_fragment"
     misc_feature    56618. .65413
                     /note="assembly_fragment"
     misc_feature    65514. .75659
                     /note="assembly_fragment"
     misc_feature    75760. .86433
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    86534. .98763
                     /note="assembly_fragment"
     misc_feature    98864. .109919
                     /note="assembly_fragment"
     misc_feature    110020. .125264
                     /note="assembly_fragment"
     misc_feature    125365. .141043
                     /note="assembly_fragment"
     misc_feature    141144. .170896
                     /note="assembly_fragment"
BASE COUNT    41547 a  41714 c  41241 g  44377 t   2017 others
ORIGIN

Alignment Scores:
Pred. No.:              1.39e-15        Length:         170896
Score:                  523.00          Matches:        99
Percent Similarity:     98.02%          Conservative:   0
Best Local Similarity:  98.02%          Mismatches:     2
Query Match:            46.41%          Indels:         0
DB:                     2               Gaps:           0

US-09-595-947E-10 (1-214) x AC011010 (1-170896)

Qy        114 ProAspAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrp 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30631 CCAGACGACGCGAAGCTCACCAAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGG 30572

Qy        134 AlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeuTyrAlaLeuGluProProAla 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30571 GCGCTGACTCAAACGCTGCGCATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCG 30512

Qy        154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTrpGlySerLeuTyr 173
              |||||||||||||||||||||||||||||||    ||||||||||||||||||||||||
Db      30511 CCGCACTGCGGGGAGCTGGGCAGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTAC 30452

Qy        174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgProGly 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      30451 TCCCCAGTCTCCCAGGCTGGCAGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGG 30392

Qy        194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerProGlySerLeuAlaPheSerAspPhe 213
              ||||||||||||||    ||||||||||||||||||||||||||||||||||||||||||
Db      30391 CTGCTGGGGGCCACCTTTTCCGCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTT 30332

Qy        214 Leu 214
              |||
Db      30331 CTG 30329


Search completed: February  2, 2004, 21:05:31
Job time : 3780 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 2, 2004, 15:49:38 ; Search time 271 Seconds
(without alignments)
2131.660 Million cell updates/sec

Title: US-09-595-947E-10
Perfect score: 1127
Sequence: 1 MTPQPSGAPTVQVTRETERS.....LGATSSACLSPGSLAFSDFL 214

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09595947/runat_02022004_154933_8036/app_query.fasta_1.391
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09595947_@CGN_1_1_0_@runat_02022004_154933_8036 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6, -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1105	98.0	5340	21	AAC61089	Human neurogenin 3
	2	1105	98.0	5340	24	AAD46871	Human neurogenin 3
	3	849	75.3	1491	19	AAV42512	cDNA encoding a no
	4	819	72.7	804	19	AAV27050	Mouse neurogenin 3
	5	819	72.7	804	21	AAZ51981	Murine neurogenin-
	6	819	72.7	861	22	AAF27266	Mouse neurogenin 3
	7	819	72.7	861	25	ABV75970	Mouse transcriptio
	8	819	72.7	1860	24	AAD46872	Murine neurogenin
	9	819	72.7	1861	21	AAC61090	Murine neurogenin
	10	819	72.7	5567	22	AAF27254	Mouse atonal homol
c	11	519	46.1	592	24	ABQ49522	Oligonucleotide fo
	12	519	46.1	592	24	ABQ49523	Oligonucleotide fo
	13	467	41.4	592	24	ABQ49524	Oligonucleotide fo
c	14	467	41.4	592	24	ABQ49525	Oligonucleotide fo
	15	388.5	34.5	790	22	AAF27264	Chicken atonal hom
	16	385.5	34.2	1074	22	AAF27263	Chicken atonal hom
	17	378.5	33.6	1268	18	AAT74891	Human neurogenic d
	18	378.5	33.6	1268	19	AAV42932	DNA encoding human
	19	378.5	33.6	1268	25	ABS56390	Human bHLH family
	20	378.5	33.6	1665	24	AAD46888	Human neurogenin 1
	21	376.5	33.4	714	24	AAD46889	Human neurogenin 1
	22	375.5	33.3	1385	19	AAV27049	Mouse neurogenin 2
	23	375.5	33.3	1385	21	AAZ51980	Murine neurogenin-
	24	375.5	33.3	1385	22	AAF27269	Mouse neurogenin 2
	25	370.5	32.9	1412	22	AAF27255	Mouse atonal homol
	26	370.5	32.9	1412	22	AAF27273	Mouse atonal homol
	27	370	32.8	1527	19	AAV27045	Rat neurogenin 1 g
	28	370	32.8	1527	21	AAZ51976	Rat neurogenin-1 (
	29	365	32.4	1332	19	AAV42938	DNA encoding murin
	30	365	32.4	1332	25	ABS56396	Mouse bHLH family
	31	365	32.4	1333	18	AAT74894	Mouse neurogenic d
	32	359.5	31.9	738	19	AAV27046	Mouse neurogenin 1
	33	359.5	31.9	738	21	AAZ51977	Murine neurogenin-
	34	346	30.7	6123	24	AAD46890	Human neurogenin 2
	35	322.5	28.6	1312	19	AAV27047	Xenopus neurogenin
	36	322.5	28.6	1312	21	AAZ51978	Xenopus X-ngnr-1a
	37	304	27.0	1277	19	AAV27048	Xenopus neurogenin
	38	304	27.0	1277	21	AAZ51979	Xenopus X-ngnr-1b
c	39	297.5	26.4	778	24	ABQ16590	Oligonucleotide fo
	40	297.5	26.4	778	24	ABQ16591	Oligonucleotide fo
	41	248	22.0	778	24	ABQ16592	Oligonucleotide fo
c	42	248	22.0	778	24	ABQ16593	Oligonucleotide fo
	43	241	21.4	2161	23	ABL13239	Drosophila melanog
c	44	241	21.4	4161	23	ABL13238	Drosophila melanog
	45	238.5	21.2	1550	22	AAF27276	Mouse atonal homol

ALIGNMENTS

RESULT 1
AAC61089
ID AAC61089 standard; DNA; 5340 BP.
XX
AC AAC61089;
XX
DT 05-FEB-2001 (first entry)
XX
DE Human neurogenin 3 (Ngn3) genomic DNA sequence.
XX
KW Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;
KW islet cell precursor identification; diabetes mellitus; human; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 3022..3666

```
FT                   /*tag=  a
FT                   /product= "Ngn3"
FT                   /note= "Neurogenin 3"
XX
PN   WO200059936-A1.
XX
PD   12-OCT-2000.
XX
PF   28-MAR-2000; 2000WO-US08436.
XX
PR   06-APR-1999;   99US-0128180.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   German MS,  Lin J;
XX
DR   WPI; 2000-664989/64.
DR   P-PSDB; AAY85617.
XX
PT   Novel human neurogenin 3 polypeptides and polynucleotides encoding
PT   them, useful for diagnosis, prevention and treatment of diabetes
PT   mellitus and to identify individuals at risk of diabetes -
XX
PS   Claim 6; Page 46-48; 54pp; English.
XX
CC   The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
CC   protein AAY85617. The Ngn3 gene is located at chromosome position
CC   10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
CC   protein sequences, and includes an antibody recognising the Ngn3 protein.
CC   Also included in the invention is a method for identifying an islet cell
CC   precursor, the method involves analysing a cell for the expression of the
CC   Ngn3 gene product, where detection of the product is indicative of an
CC   islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
CC   reagent for detecting (in a subject) a predisposition to a defect in
CC   pancreatic islet cell function or formation associated with a defect in
CC   Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
CC   precursor cells expressing Ngn3, and to alter cellular differentiation in
CC   culture in vivo to produce new beta-cells to treat patients with diabetes
CC   mellitus.
XX
SQ   Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

Alignment Scores:
Pred. No.:              1.65e-59        Length:         5340
Score:                  1105.00         Matches:        211
Percent Similarity:     99.07%          Conservative:   1
Best Local Similarity:  98.60%          Mismatches:     2
Query Match:            98.05%          Indels:         0
DB:                     21              Gaps:           0

US-09-595-947E-10 (1-214) x AAC61089 (1-5340)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3022 ATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGAGCGGTCC 3081

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3082 TTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCCAGCCCCACT 3141

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3142 CGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCCCCGAGGAAG 3201

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3202 CTCCGGGCACGGCGCGGGGGACGCAGCCGGCCTAAGAGCGAGTTGGCACTGAGCAAGCAG 3261

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       3262 CGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACAACCTCAAC 3321

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3322 TCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCGAAGCTCACC 3381

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3382 AAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAACGCTGCGC 3441

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3442 ATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGAGCTGGGC 3501

Qy        161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
              ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db       3502 AGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCAGGCTGGC 3561

Qy        181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3562 AGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCACCTCTTCC 3621

Qy        201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
              ||||||||||||||||||||||||||||||||||||||||||
Db       3622 GCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTG 3663

RESULT 2
AAD46871
ID   AAD46871 standard; DNA; 5340 BP.
XX
AC   AAD46871;
XX
DT   27-JAN-2003  (first entry)
XX
DE   Human neurogenin 3 (Ngn3) gene.
XX
KW   Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW   type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW   islet cell; cell therapy; neurogenin 3; Ngn3; chromosome 10q22.1-22.2;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             3022..3666
FT                   /*tag=  a
FT                   /product= "Human Ngn3 protein"
XX
PN   WO200274045-A2.
XX
PD   26-SEP-2002.
XX
PF   20-MAR-2002; 2002WO-US11166.
XX
PR   20-MAR-2001; 2001US-0817360.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   German MS,  Lin J;
XX
DR   WPI; 2002-759853/82.
DR   P-PSDB; AAE29277.
XX
PT   Producing a mammalian islet cell for treating diabetes mellitus
PT   comprises introducing into a mammalian cell a nucleic acid molecule
PT   encoding neuroendocrine basic helix-loop-helix transcription factor  -
XX
PS   Example 2; Page 87-88; 108pp; English.
XX
CC   The invention relates to a method for producing a mammalian islet cell.
CC   The method comprising introducing into a mammalian cell a nucleic acid
CC   molecule encoding an islet transcription factor for expression of the
CC   islet transcription factor in the cell and for production of islet cell
CC   phenotype in the cell. The islet transcription factor is a neuroendocrine
CC   basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC   for treating type 2 diabetes mellitus and for replacing beta cells lost
```

```
CC   to autoimmune destruction in individuals with type 1 diabetes. The method
CC   is useful in cell therapy. The present sequence is human neurogenin 3
CC   (Ngn3) gene. Ngn3 gene is located on chromosome 10q22.1-22.2.
XX
SQ   Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

Alignment Scores:
Pred. No.:              1.65e-59      Length:        5340
Score:                  1105.00       Matches:       211
Percent Similarity:     99.07%        Conservative:  1
Best Local Similarity:  98.60%        Mismatches:    2
Query Match:            98.05%        Indels:        0
DB:                     24            Gaps:          0

US-09-595-947E-10 (1-214) x AAD46871 (1-5340)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3022 ATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGAGCGGTCC 3081

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3082 TTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCCAGCCCCACT 3141

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3142 CGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCCCCGAGGAAG 3201

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3202 CTCCGGGCACGGCGCGGGGGACGCAGCCGGCCTAAGAGCGAGTTGGCACTGAGCAAGCAG 3261

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       3262 CGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACAACCTCAAC 3321

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3322 TCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCGAAGCTCACC 3381

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3382 AAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAACGCTGCGC 3441

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3442 ATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGAGCTGGGC 3501

Qy        161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
              ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db       3502 AGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCAGGCTGGC 3561

Qy        181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3562 AGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCACCTCTTCC 3621

Qy        201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
              |||||||||||||||||||||||||||||||||||||||||||
Db       3622 GCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTG 3663

RESULT 3
AAV42512
ID   AAV42512 standard; cDNA; 1491 BP.
XX
AC   AAV42512;
XX
DT   05-OCT-1998  (first entry)
XX
DE   cDNA encoding a novel BHLH protein designated RELAX.
XX
KW   Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;
KW   control; gene expression; transcriptional activator; targeting;
KW   protein expression; central nervous system; CNS; treatment;
KW   nervous system disorder; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             459..1103
FT                   /*tag= a
FT                   /product= RELAX
XX
PN   WO9827206-A2.
XX
PD   25-JUN-1998.
XX
PF   19-DEC-1997;   97WO-FR02368.
XX
PR   19-DEC-1996;   96FR-0015651.
XX
PA   (RHON ) RHONE-POULENC RORER SA.
XX
PI   Mallet J,  Ravassard P,  Icard-Liepkalns C;
XX
DR   WPI; 1998-362775/31.
DR   P-PSDB; AAW62991.
XX
PT   Basic helix-loop-helix polypeptide and related nucleic acid - with
PT   transcriptional activity, for targeting expression of genes to
PT   central nervous system and treatment of nervous disease
XX
PS   Claim 6; Page 20; 28pp; French.
XX
CC   The present sequence encodes a basic helix-loop-helix (BHLH) type
CC   protein, designated RELAX (Rat Embryonic Longitudinal Axis) protein.
CC   The protein is used to control and  participate in gene expression,
CC   by acting as transcriptional activator, strictly dependent on the
CC   presence of an intact E box (CANNTG), particularly for targeting
CC   expression of proteins to the central nervous system (CNS). The
CC   nucleic acid sequence can be used to treat nervous system disorders,
CC   and antisense sequences can be used to control mRNA transcription.
XX
SQ   Sequence 1491 BP; 307 A; 487 C; 413 G; 284 T; 0 other;

Alignment Scores:
Pred. No.:              3.5e-44       Length:        1491
Score:                  849.00        Matches:       166
Percent Similarity:     83.72%        Conservative:  14
Best Local Similarity:  77.21%        Mismatches:    33
Query Match:            75.33%        Indels:        2
DB:                     19            Gaps:          2

US-09-595-947E-10 (1-214) x AAV42512 (1-1491)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||      |||||||||:::||||||::::::||||||::::::
Db        459 ATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCC 518

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||  ||||||:::  ||||||           |||  ||||||||||||||||
Db        519 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 578

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                 |||   :::|||:::|||||||||   |||   |||||||||      ||||||
Db        579 CTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAG 638

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        639 CTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAG 698

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        699 CGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAAC 758

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        759 TCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 818

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        819 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGC 878

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||| |||      |||||||| |||  ||||||||||||||||
Db        879 ATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC---TGTGGGGAGCTGGGA 935

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||     ||||||||||||||:::||||||||||||||||||||
Db        936 AGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCT 995

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              ||||||||||||||   ||||||||||||||   |||||||||      :::|||
Db        996 GGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCC 1055

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                :::||||||   ||||||:::|||   ||||||||||||||
Db       1056 CCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTG 1100

RESULT 4
AAV27050
ID   AAV27050 standard; cDNA; 804 BP.
XX
AC   AAV27050;
XX
DT   17-SEP-1998  (first entry)
XX
DE   Mouse neurogenin 3 gene.
XX
KW   ds; Mouse; neurogenin; expression vector; recombinant protein;
KW   antibody; neurogenesis.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             160..804
FT                   /*tag=  a
FT                   /product= "Mouse neurogenin 3"
XX
PN   WO9813491-A2.
XX
PD   02-APR-1998.
XX
PF   24-SEP-1997;   97WO-US17048.
XX
PR   17-SEP-1997;   97US-0932411.
PR   27-SEP-1996;   96US-0722570.
PR   12-NOV-1996;   96US-0030864.
PR   19-DEC-1996;   96US-0772009.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Anderson DJ,  Ma Q,  Sommer L;
XX
DR   WPI; 1998-230702/20.
DR   P-PSDB; AAW54947.
XX
PT   Mouse neurogenins, useful in neurogenesis - and recombinant nucleic
PT   acids and proteins derived from rat and Xenopus
XX
PS   Disclosure; Fig 9; 106pp; English.
XX
CC   The Mouse neurogenin 3 is one of several neurogenin proteins discussed
CC   in the present invention.  The neurogenin nucleic acids can be expressed
CC   in a host cell, transformed using an expression vector, to produce
CC   recombinant proteins.  The proteins and the antibodies raised against
CC   the proteins are useful in the study of neurogenesis.
XX
SQ   Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;
```

```
Alignment Scores:
Pred. No.:              1.35e-42        Length:         804
Score:                  819.00          Matches:        163
Percent Similarity:     82.33%          Conservative:   14
Best Local Similarity:  75.81%          Mismatches:     36
Query Match:            72.67%          Indels:         2
DB:                     19              Gaps:           2

US-09-595-947E-10 (1-214) x AAV27050 (1-804)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||      |||   |||:::||||||:::   ||||||::::::
Db        160 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 219

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||          |||   ||||||||||||||
Db        220 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 279

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                  |||   :::|||:::|||||||||   |||   |||||||||       ||||||
Db        280 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 339

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db        340 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 399

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        400 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 459

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 519

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 579

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||| |||      |||||||| |||  ||||||||||||||||
Db        580 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 636

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||     ||||||||||||||:::||||||||||||||||||||
Db        637 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 696

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   ||||||||||||||   |||||||||      :::|||
Db        697 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 756

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                :::   |||   ||||||:::|||   ||||||||||||||
Db        757 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 801

RESULT 5
AAZ51981
ID   AAZ51981 standard; DNA; 804 BP.
XX
AC   AAZ51981;
XX
DT   04-JUL-2000  (first entry)
XX
DE   Murine neurogenin-3 (NGN3) nucleic acid sequence.
XX
KW   Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;
KW   Phox2a protein; neuronal subtype-specific marker; growth factor;
KW   neural differentiation; transplantation; neuronal dysfunction;
KW   optical nerve damage; auditory nerve damage, neurodegenerative disorder;
KW   neuroprotective; nootropic; anticonvulsant; antiParkinsonian; vulnerary;
KW   cerebroprotective; immunesuppressant; antiinfectious; ds.
XX
OS   Mus sp.
XX
```

```
FH   Key             Location/Qualifiers
FT   CDS             160..804
FT                   /*tag= a
FT                   /product= "Murine neurogenin-3 protein"
XX
PN   WO200009676-A2.
XX
PD   24-FEB-2000.
XX
PF   13-AUG-1999;   99WO-US18525.
XX
PR   14-AUG-1998;   98US-0096630.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Anderson DJ,  Lo L;
XX
DR   WPI; 2000-256250/22.
DR   P-PSDB; AAY70570.
XX
PT   Inducing non-neuronal cells to differentiate into neurons and for
PT   non-neuronal cells to express a neuronal subtype-specific marker,
PT   comprising contacting the non-neuronal cells with a vector containing
PT   neurogenin nucleic acid -
XX
PS   Claim 1; Fig 1J; 76pp; English.
XX
CC   The patent discloses a method for inducing non-neuronal cells (NNC) to
CC   differentiate into neurons and for NNCs to express a neuronal subtype
CC   -specific marker. Transformed host cells are used as sources of neuronal
CC   and other growth factors; in culture for screening compounds that
CC   modulate neural differentiation or as sources of recombinantly produced
CC   neurogenins and Phox2a proteins for use in transplantation. The cells
CC   also have a variety of in vivo uses, e.g. for transplantation at sites of
CC   neuronal dysfunction e.g. patients with hearing or vision loss due to
CC   optical or auditory nerve damage, brain or spinal cord injuries, and
CC   neurodegenerative disorders e.g. Alzheimer's disease. The present
CC   sequence encodes murine neurogenin-3 (NGN-3), a transcription factor.
CC   NNCs differentiate into neurons through the recombinant expression of a
CC   transcription factor that induces a core program of neurogenesis. Forced
CC   expression of murine NGN3 can elicit expression of at least some neuronal
CC   phenotypic markers even in NNCs.
XX
SQ   Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

Alignment Scores:
Pred. No.:              1.35e-42        Length:         804
Score:                  819.00          Matches:        163
Percent Similarity:     82.33%          Conservative:   14
Best Local Similarity:  75.81%          Mismatches:     36
Query Match:            72.67%          Indels:         2
DB:                     21              Gaps:           2

US-09-595-947E-10 (1-214) x AAZ51981 (1-804)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::|||||||:::   ||||||::::::
Db        160 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 219

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||            |||   ||||||||||||||
Db        220 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 279

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                  |||   :::|||:::||||||||||   |||   |||||||||      ||||||
Db        280 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 339

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db        340 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 399

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||

Db        400 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 459

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 519

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 579

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||       |||||||||   |||   ||||||||||||||
Db        580 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 636

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::||||||||||||||||||||
Db        637 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 696

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   ||||||||||||||   |||||||||         :::|||
Db        697 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 756

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                  :::   |||   ||||||:::|||   ||||||||||||||
Db        757 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 801

RESULT 6
AAF27266
ID   AAF27266 standard; cDNA; 861 BP.
XX
AC   AAF27266;
XX
DT   24-APR-2001  (first entry)
XX
DE   Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.
XX
KW   Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW   hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW   cellular proliferation; cerebellar granule neuron; gene therapy;
KW   mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW   transgenic animal; ss.
XX
OS   Mus musculus.
XX
PN   WO200073764-A2.
XX
PD   07-DEC-2000.
XX
PF   01-JUN-2000; 2000WO-US15410.
XX
PR   01-JUN-1999;   99US-0137060.
PR   19-JAN-2000; 2000US-0176993.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Zoghbi HY,  Bellen H,  Birmingham N,  Hassan B,  Ben-Arie N;
XX
DR   WPI; 2001-032190/04.
DR   P-PSDB; AAB60359.
XX
PT   Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT   any of its homologs or orthologs, for the treatment of e.g. deafness,
PT   osteoarthritis and abnormal cell proliferation -
XX
PS   Disclosure; Page -; 142pp; English.
XX
CC   The invention relates to the use of atonal-associated nucleic acid or
CC   amino acid sequence, or any of its homologues or orthologues as
CC   therapeutic agents for the treatment of deafness, partial hearing loss,
CC   vestibular effects due to damage or loss of inner hair cells,
CC   osteoarthritis and abnormal cell proliferation. The invention also
CC   encompasses methods of screening for compounds which affect the
CC   expression of an atonal-associated nucleic acid sequence in an animal,
```

```
CC   and a transgenic animal in which an allele of a native atonal-associated
CC   gene is replaced by a heterologous nucleic acid sequence, thus
CC   inactivating the atonal-associated allele. The nucleic acids or proteins
CC   may be used in a method of treating an animal for hearing impairment,
CC   joint disease, balance disorders, abnormal cell proliferation, or other
CC   disease related to loss of a functional atonal-associated nucleic acid or
CC   protein. They may particularly be used to treat an animal with a
CC   deficiency in cerebellar granule neurons or their precursors, and may
CC   also be used in promoting mechanoreceptive cell growth and generating
CC   hair cells. The present sequence represents an atonal-associated nucleic
CC   acid sequence referred to in the invention.
CC   Note: The present sequence is not shown in the specification, but
CC   was obtained from GenBank.
XX
SQ   Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

Alignment Scores:
Pred. No.:              1.45e-42      Length:         861
Score:                  819.00        Matches:        163
Percent Similarity:     82.33%        Conservative:   14
Best Local Similarity:  75.81%        Mismatches:     36
Query Match:            72.67%        Indels:         2
DB:                     22            Gaps:           2

US-09-595-947E-10 (1-214) x AAF27266 (1-861)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::||||||:::   ||||||::::::
Db        160 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 219

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||          |||   ||||||||||||||||
Db        220 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 279

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                    |||   :::|||:::||||||||||   |||   |||||||||       ||||||
Db        280 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 339

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        340 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 399

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        400 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 459

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 519

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 579

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||       |||||||||   |||   ||||||||||||||
Db        580 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 636

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||       ||||||||||||||:::||||||||||||||||||||
Db        637 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 696

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   ||||||||||||||   |||||||||       :::|||
Db        697 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 756

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
              :::   |||   ||||||:::|||   ||||||||||||||
Db        757 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 801

RESULT 7
ABV75970
ID   ABV75970 standard; cDNA; 861 BP.
XX
AC   ABV75970;
XX
DT   11-FEB-2003  (first entry)
XX
DE   Mouse transcription factor neurogenin 3 cDNA.
XX
KW   Mouse; transcription factor; neurogenin 3; ngn3; stem cell;
KW   differentiation; beta-cell; insulin; diabetes; hyperglycaemia;
KW   glucose intolerance; antidiabetic; hypoglycaemic; gene therapy;
KW   gene; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             160..804
FT                   /*tag=  a
FT                   /product= "Mouse ngn3"
XX
PN   WO200286107-A2.
XX
PD   31-OCT-2002.
XX
PF   19-APR-2002; 2002WO-EP04362.
XX
PR   19-APR-2001; 2001US-284531P.
XX
PA   (DEVE-) DEVELOGEN ENTWICKLUNGSBIOLOGISCHE FORSCH.
PA   (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.
XX
PI   Wobus AM,  St-Onge L,  Blyszczuk P,  Hoffmann U;
XX
DR   WPI; 2003-075629/07.
XX
PT   Differentiating stem cells into insulin-producing cells useful for
PT   treating pancreatic diseases, by culturing stem cells in suitable
PT   medium and activating gene involved in beta-cell differentiation -
XX
PS   Disclosure; Page 58-59; 62pp; English.
XX
CC   The present sequence is that of cDNA encoding the murine
CC   basic helix-loop-helix transcription factor neurogenin 3 (ngn3),
CC   a gene which is required for the specification of the early
CC   endocrine precursor in the pancreatic epithelium and which is
CC   down-regulated once endocrine differentiation begins. The invention
CC   provides a claimed method for differentiating stem cells (especially
CC   embryonic, adult or somatic stem cells and primordial germ cells)
CC   into insulin-producing cells. This involves culturing stem cells in
CC   a suitable medium and activating at least one gene involved in
CC   beta-cell differentiation. Preferred genes including Pdx1, Pax4,
CC   Pax6 and ngn3 (see ABV75967-70). Gene activation comprises the
CC   delivery of the gene into stem cells using a viral delivery
CC   system, or the delivery of a protein product of the gene into stem
CC   cells. The insulin-producing cells can be transplanted into
CC   animals or human for treatment of pancreatic diseases, metabolic
CC   syndrome and metabolic disorders with impaired glucose levels such
CC   as diabetes, hyperglycaemia and impaired glucose tolerance
CC   (claimed). The cells can also be used to identify compounds which
CC   stimulate beta-cell differentiation, insulin secretion or glucose
CC   responsiveness. Differentiated beta-cells can be used to study the
CC   toxic and other effects of exogenous compounds on beta-cell
CC   function. In an example from the invention, Pax6 cDNA was inserted
CC   into expression vector pACCMV.pLpA under the control of the
CC   cytomegalovirus promoter.
XX
SQ   Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

Alignment Scores:
Pred. No.:              1.45e-42      Length:         861
Score:                  819.00        Matches:        163
Percent Similarity:     82.33%        Conservative:   14
Best Local Similarity:  75.81%        Mismatches:     36
Query Match:            72.67%        Indels:         2
```

```
DB:                    25            Gaps:          2

US-09-595-947E-10 (1-214) x ABV75970 (1-861)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::||||||:::   ||||||::::::
Db        160 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 219

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||           |||   ||||||||||||||
Db        220 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 279

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                  |||   :::|||:::||||||||   |||   |||||||||      ||||||
Db        280 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 339

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        340 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 399

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        400 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 459

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 519

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 579

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||      |||||||||   |||   ||||||||||||||
Db        580 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 636

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::|||||||||||||||||||
Db        637 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 696

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::||||||||||   ||||||||||||||   |||||||||        :::|||
Db        697 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 756

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                 :::   |||   ||||||:::|||   ||||||||||||||
Db        757 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 801

RESULT 8
AAD46872
ID   AAD46872 standard; DNA; 1860 BP.
XX
AC   AAD46872;
XX
DT   27-JAN-2003  (first entry)
XX
DE   Murine neurogenin 3 (Ngn3) gene.
XX
KW   Transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW   type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW   islet cell; cell therapy; neurogenin 3; Ngn3; murine; gene; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1093..1737
FT                   /*tag=  a
FT                   /product= "Murine Ngn3 protein"
XX
PN   WO200274045-A2.
XX
PD   26-SEP-2002.
XX
PF   20-MAR-2002; 2002WO-US11166.
XX
PR   20-MAR-2001; 2001US-0817360.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   German MS,  Lin J;
XX
DR   WPI; 2002-759853/82.
DR   P-PSDB; AAE29278.
XX
PT   Producing a mammalian islet cell for treating diabetes mellitus
PT   comprises introducing into a mammalian cell a nucleic acid molecule
PT   encoding neuroendocrine basic helix-loop-helix transcription factor  -
XX
PS   Example 3; Page 89-90; 108pp; English.
XX
CC   The invention relates to a method for producing a mammalian islet cell.
CC   The method comprising introducing into a mammalian cell a nucleic acid
CC   molecule encoding an islet transcription factor for expression of the
CC   islet transcription factor in the cell and for production of islet cell
CC   phenotype in the cell. The islet transcription factor is a neuroendocrine
CC   basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC   for treating type 2 diabetes mellitus and for replacing beta cells lost
CC   to autoimmune destruction in individuals with type 1 diabetes. The method
CC   is useful in cell therapy. The present sequence is murine neurogenin 3
CC   (Ngn3) gene.
XX
SQ   Sequence 1860 BP; 397 A; 559 C; 537 G; 367 T; 0 other;

Alignment Scores:
Pred. No.:              3.21e-42        Length:         1860
Score:                  819.00          Matches:        163
Percent Similarity:     82.33%          Conservative:   14
Best Local Similarity:  75.81%          Mismatches:     36
Query Match:            72.67%          Indels:         2
DB:                     24              Gaps:           2

US-09-595-947E-10 (1-214) x AAD46872 (1-1860)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::||||||:::   ||||||::::::
Db       1093 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 1152

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||           |||   ||||||||||||||
Db       1153 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 1212

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                  |||   :::|||:::||||||||   |||   |||||||||      ||||||
Db       1213 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 1272

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db       1273 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 1332

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       1333 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 1392

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1393 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 1452

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1453 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 1512

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||      |||||||||   |||   ||||||||||||||
Db       1513 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 1569

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
```

```
Db       1570 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 1629

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199

Db       1630 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 1689

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214

Db       1690 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 1734
```

RESULT 9
AAC61090
ID AAC61090 standard; DNA; 1861 BP.
XX
AC AAC61090;
XX
DT 05-FEB-2001 (first entry)
XX
DE Murine neurogenin 3 (Ngn3) genomic DNA sequence.
XX
KW Neurogenin 3; Ngn3; cellular differentiation; diabetes mellitus;
KW islet cell precursor identification; mouse; ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 1093..1737
FT /*tag= a
FT /product= "Ngn3"
FT /note= "Neurogenin 3"
XX
PN WO200059936-A1.
XX
PD 12-OCT-2000.
XX
PF 28-MAR-2000; 2000WO-US08436.
XX
PR 06-APR-1999; 99US-0128180.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI German MS, Lin J;
XX
DR WPI; 2000-664989/64.
DR P-PSDB; AAY85618.
XX
PT Novel human neurogenin 3 polypeptides and polynucleotides encoding
PT them, useful for diagnosis, prevention and treatment of diabetes
PT mellitus and to identify individuals at risk of diabetes -
XX
PS Claim 18; Page 49-50; 54pp; English.
XX
CC The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
CC protein AAY85617. The Ngn3 gene is located at chromosome position
CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
CC protein sequences, and includes an antibody recognising the Ngn3 protein.
CC Also included in the invention is a method for identifying an islet cell
CC precursor, the method involves analysing a cell for the expression of the
CC Ngn3 gene product, where detection of the product is indicative of an
CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
CC reagent for detecting (in a subject) a predisposition to a defect in
CC pancreatic islet cell function or formation associated with a defect in
CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
CC precursor cells expressing Ngn3, and to alter cellular differentiation in
CC culture in vivo to produce new beta-cells to treat patients with diabetes
CC mellitus. The present sequence represents the murine Ngn3 genomic DNA
CC sequence.
XX
SQ Sequence 1861 BP; 397 A; 560 C; 537 G; 367 T; 0 other;

Alignment Scores:
Pred. No.: 3.21e-42 Length: 1861

Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 21 Gaps: 2

US-09-595-947E-10 (1-214) x AAC61090 (1-1861)

```
Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20

Db       1093 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 1152

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40

Db       1153 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 1212

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60

Db       1213 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 1272

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80

Db       1273 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 1332

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100

Db       1333 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 1392

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrProAspAspAlaLysLeuThr 120

Db       1393 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 1452

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140

Db       1453 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 1512

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160

Db       1513 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 1569

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179

Db       1570 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 1629

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199

Db       1630 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 1689

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214

Db       1690 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 1734
```

RESULT 10
AAF27254
ID AAF27254 standard; cDNA; 5567 BP.
XX
AC AAF27254;
XX
DT 24-APR-2001 (first entry)
XX
DE Mouse atonal homologue 5 (ATOH5, Math4B) cDNA, SEQ ID NO:4.
XX
KW Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW cellular proliferation; cerebellar granule neuron; gene therapy;
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW transgenic animal; ss.
XX
OS Mus musculus.
XX
PN WO200073764-A2.
XX
PD 07-DEC-2000.

XX
PF 01-JUN-2000; 2000WO-US15410.
XX
PR 01-JUN-1999; 99US-0137060.
PR 19-JAN-2000; 2000US-0176993.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
XX
DR WPI; 2001-032190/04.
DR P-PSDB; AAB60350.
XX
PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT any of its homologs or orthologs, for the treatment of e.g. deafness,
PT osteoarthritis and abnormal cell proliferation -
XX
PS Disclosure; Page -; 142pp; English.
XX
CC The invention relates to the use of atonal-associated nucleic acid or
CC amino acid sequence, or any of its homologues or orthologues as
CC therapeutic agents for the treatment of deafness, partial hearing loss,
CC vestibular effects due to damage or loss of inner hair cells,
CC osteoarthritis and abnormal cell proliferation. The invention also
CC encompasses methods of screening for compounds which affect the
CC expression of an atonal-associated nucleic acid sequence in an animal,
CC and a transgenic animal in which an allele of a native atonal-associated
CC gene is replaced by a heterologous nucleic acid sequence, thus
CC inactivating the atonal-associated allele. The nucleic acids or proteins
CC may be used in a method of treating an animal for hearing impairment,
CC joint disease, balance disorders, abnormal cell proliferation, or other
CC disease related to loss of a functional atonal-associated nucleic acid or
CC protein. They may particularly be used to treat an animal with a
CC deficiency in cerebellar granule neurons or their precursors, and may
CC also be used in promoting mechanoreceptive cell growth and generating
CC hair cells. The present sequence represents an atonal-associated nucleic
CC acid sequence referred to in the invention.
CC Note: The present sequence is not shown in the specification, but
CC was obtained from GenBank.
XX
SQ Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;

Alignment Scores:

Pred. No.:	9.91e-42	Length:	5567
Score:	819.00	Matches:	163
Percent Similarity:	82.33%	Conservative:	14
Best Local Similarity:	75.81%	Mismatches:	36
Query Match:	72.67%	Indels:	2
DB:	22	Gaps:	2

US-09-595-947E-10 (1-214) x AAF27254 (1-5567)

```
Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::|||||:::   ||||||::::::
Db       4923 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 4982

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||           |||   ||||||||||||||||
Db       4983 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 5042

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                  |||   :::|||:::|||||||||   |||   |||||||||       ||||||
Db       5043 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 5102

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db       5103 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 5162

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       5163 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 5222

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5223 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 5282

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5283 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 5342

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||       |||||||||   |||   ||||||||||||||
Db       5343 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 5399

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||       ||||||||||||||:::||||||||||||||||||||
Db       5400 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 5459

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   |||||||||||||||   |||||||||          :::|||
Db       5460 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 5519

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                 :::   |||   ||||||:::|||   ||||||||||||||
Db       5520 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 5564
```

RESULT 11
ABQ49522/c
ID ABQ49522 standard; DNA; 592 BP.
XX
AC ABQ49522;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36113.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

```
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 592 BP; 81 A; 59 C; 201 G; 251 T; 0 other;

Alignment Scores:
Pred. No.:              4.21e-24        Length:         592
Score:                  519.00          Matches:        98
Percent Similarity:     83.67%          Conservative:   25
Best Local Similarity:  66.67%          Mismatches:     24
Query Match:            46.05%          Indels:         0
DB:                     24              Gaps:           0

US-09-595-947E-10 (1-214) x ABQ49522 (1-592)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              :::|||||||||||||||||     ||||||:::|||:::||||||:::|||:::||||||
Db        443 ATAACGCCTCAACCCTCGAATACGCCCACTATCCAAATAACCCGTAAAACGAAACGATCC 384

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||:::   |||::::::|||:::|||   |||||||||||||||||||:::||||||
Db        383 TTCCCCAAAACCTCGAAAAACGAAATAACCTACCCCACGTCCGCCCCGCCCAACCCCACT 324

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
              ||||||      |||   |||:::   ::::::          |||       |||:::|||
Db        323 CGCACACGAAAAAACTACGCAAAAACGAAAAAAAAAAACTACCGAAAAACCCCGAAAAAA 264

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||   ||||||      |||:::|||   |||:::||||||   |||:::||||||
Db        263 CTCCGAACACGACGCGAAAAACGCAACCGACTTAAAAACGAATTAACACTAAACAAACAA 204

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||:::||||||||||||   ||||||||||||||||||:::|||:::||||||
Db        203 CGACGAAATCGACGAAAAAAAAACCAACGACCGCGAACGCAATCGAATACACAACCTCAAC 144

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              |||   |||:::|||||||||   :::|||||||||||||||:::|||||||||||||||
Db        143 TCGACACTAAACGCCCTACGCGATATCCTACCCACCTTCCCAAACGACGCGAAACTCACC 84

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db         83 AAAATCGAAACGCTACGCTTCGCCCACAACTACATCTAAACGCTAACTCAAACGCTACGC 24

Qy        141 IleAlaAspHisSerLeuTyr 147
              |||   :::|||:::||||||
Db         23 ATAACGAACCACAACTTATAC 3

RESULT 12
ABQ49523
ID   ABQ49523 standard; DNA; 592 BP.
XX
AC   ABQ49523;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 36114.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;

Alignment Scores:
Pred. No.:              4.21e-24        Length:         592
Score:                  519.00          Matches:        98
Percent Similarity:     83.67%          Conservative:   25
Best Local Similarity:  66.67%          Mismatches:     24
Query Match:            46.05%          Indels:         0
DB:                     24              Gaps:           0

US-09-595-947E-10 (1-214) x ABQ49523 (1-592)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              :::|||||||||||||||||     ||||||:::|||:::||||||:::|||:::||||||
Db        150 ATAACGCCTCAACCCTCGAATACGCCCACTATCCAAATAACCCGTAAAACGAAACGATCC 209

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||:::   |||::::::|||:::|||   |||||||||||||||||||:::||||||
Db        210 TTCCCCAAAACCTCGAAAAACGAAATAACCTACCCCACGTCCGCCCCGCCCAACCCCACT 269

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
              ||||||      |||   |||:::   ::::::          |||       |||:::|||
Db        270 CGCACACGAAAAAACTACGCAAAAACGAAAAAAAAAAACTACCGAAAAACCCCGAAAAAA 329

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||   ||||||      |||:::|||   |||:::||||||   |||:::||||||
Db        330 CTCCGAACACGACGCGAAAAACGCAACCGACTTAAAAACGAATTAACACTAAACAAACAA 389

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||:::||||||||||||   ||||||||||||||||||:::|||:::||||||
Db        390 CGACGAAATCGACGAAAAAAAAACCAACGACCGCGAACGCAATCGAATACACAACCTCAAC 449

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              |||   |||:::||||||||||   :::|||||||||||||||:::|||||||||||||||
```

```
Db        450 TCGACACTAAACGCCCTACGCGATATCCTACCCACCTTCCCAAACGACGCGAAACTCACC 509

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||      ||||||||||||||||||
Db        510 AAAATCGAAACGCTACGCTTCGCCCACAACTACATCTAAACGCTAACTCAAACGCTACGC 569

Qy        141 IleAlaAspHisSerLeuTyr 147
              |||   :::|||:::||||||
Db        570 ATAACGAACCACAACTTATAC 590
```

RESULT 13
ABQ49524
ID ABQ49524 standard; DNA; 592 BP.
XX
AC ABQ49524;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36115.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 592 BP; 123 A; 59 C; 187 G; 223 T; 0 other;

Alignment Scores:
Pred. No.: 7.13e-21 Length: 592

```
Score:                  467.00          Matches:        101
Percent Similarity:     73.94%          Conservative:   4
Best Local Similarity:  71.13%          Mismatches:     37
Query Match:            41.44%          Indels:         0
DB:                     24              Gaps:           0

US-09-595-947E-10 (1-214) x ABQ49524 (1-592)

Qy          6 SerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPheProArgAlaSer 25
              |||||||||     |||   |||   ||||||||||||||   |||   |||   |||
Db        165 TCGGGTGCGTTTATTGTTTAAGTGATTCGTGAGACGGAGCGGTTTTTTTTTAGAGTTTCG 224

Qy         26 GluAspGluValThrCysProThrSerAlaProProSerProThrArgThrProGlyAsn 45
              ||||||||||||   |||   |||             |||       |||       ||||||
Db        225 GAAGACGAAGTGATTTGTTTTACGTTCGTTTCGTTTAGTTTTATTCGTATACGGGGGAAT 284

Qy         46 CysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLysLeuArgAlaArgArg 65
              |||   ||||||||||||||||||||||||||       ||||||   |||   ||||||
Db        285 TGCGTAGAGGCGGAAGAGGGAGGTTGTCGAGGGGTTTCGAGGAAGTTTCGGGTACGGCGC 344

Qy         66 GlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArg 85
              ||||||||||||||   ||||||||||   |||||||||   ||||||||||||||||
Db        345 GGGGGACGTAGTCGGTTTAAGAGCGAGTTGGTATTGAGTAAGTAGCGACGGAGTCGGCGA 404

Qy         86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
              ||||||   ||||||||||||||||||||||:::::   ||||||   ||||||
Db        405 AAGAAGGTTAACGATCGCGAGCGTAATCGAATGTATAATTTTAATTCGGTATTGGACGTT 464

Qy        106 LeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeu 125
              |||||||||||||||       |||   ||||||||||       ||||||||||||||||
Db        465 TTGCGCGGTGTTTTGTTTATTTTTTAGACGACGCGAAGTTTATTAAGATCGAGACGTTG 524

Qy        126 ArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSer 145
              ||||||   :::|||||||||||||||||||      |||||||||||||||||:::|||
Db        525 CGTTTCGTTTATAATTATATTTGGGCGTTGATTTAAACGTTGCGTATAGCGGATTATAGT 584

Qy        146 LeuTyr 147
              ||||||
Db        585 TTGTAC 590
```

RESULT 14
ABQ49525/c
ID ABQ49525 standard; DNA; 592 BP.
XX
AC ABQ49525;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36116.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful

```
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 592 BP; 223 A; 187 C; 59 G; 123 T; 0 other;

Alignment Scores:
Pred. No.:              7.13e-21        Length:         592
Score:                  467.00          Matches:        101
Percent Similarity:     73.94%          Conservative:   4
Best Local Similarity:  71.13%          Mismatches:     37
Query Match:            41.44%          Indels:         0
DB:                     24              Gaps:           0

US-09-595-947E-10 (1-214) x ABQ49525 (1-592)

Qy          6 SerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPheProArgAlaSer 25
              |||||||||      |||   |||   ||||||||||||||||   |||   |||   |||
Db        428 TCGGGTGCGTTTATTGTTTAAGTGATTCGTGAGACGGAGCGGTTTTTTTTTAGAGTTTCG 369

Qy         26 GluAspGluValThrCysProThrSerAlaProProSerProThrArgThrProGlyAsn 45
              ||||||||||||   |||   |||              |||       |||       ||||||
Db        368 GAAGACGAAGTGATTTGTTTTACGTTCGTTTCGTTTAGTTTTATTCGTATACGGGGGAAT 309

Qy         46 CysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLysLeuArgAlaArgArg 65
              |||   ||||||||||||||||||||||||||      ||||||   |||   ||||||
Db        308 TGCGTAGAGGCGGAAGAGGGAGGTTGTCGAGGGGTTTCGAGGAAGTTTCGGGTACGGCGC 249

Qy         66 GlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArg 85
              ||||||||||||||   ||||||||||||   |||||||||   ||||||||||||||||
Db        248 GGGGGACGTAGTCGGTTTAAGAGCGAGTTGGTATTGAGTAAGTAGCGACGGAGTCGGCGA 189

Qy         86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
              ||||||   ||||||||||||||||||||||:::::   ||||||   ||||||
Db        188 AAGAAGGTTAACGATCGCGAGCGTAATCGAATGTATAATTTTAATTCGGTATTGGACGTT 129

Qy        106 LeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeu 125
              ||||||||||||||      |||   ||||||||||||      ||||||||||||||||
Db        128 TTGCGCGGTGTTTTGTTTATTTTTTTAGACGACGCGAAGTTTATTAAGATCGAGACGTTG 69

Qy        126 ArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSer 145
              ||||||   :::|||||||||||||||||      ||||||||||||||:::|||
Db         68 CGTTTCGTTTATAATTATATTTGGGCGTTGATTTAAACGTTGCGTATAGCGGATTATAGT 9

Qy        146 LeuTyr 147
              ||||||
Db          8 TTGTAC 3

RESULT 15
AAF27264
ID   AAF27264 standard; cDNA; 790 BP.
XX
AC   AAF27264;
XX
DT   24-APR-2001  (first entry)
XX
DE   Chicken atonal homologue ngn2/ath4a cDNA, SEQ ID NO:20.
XX
KW   Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW   hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW   cellular proliferation; cerebellar granule neuron; gene therapy;
KW   mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW   transgenic animal; ss.
XX
OS   Gallus gallus.
XX
PN   WO200073764-A2.
XX
PD   07-DEC-2000.
XX
PF   01-JUN-2000; 2000WO-US15410.
XX
PR   01-JUN-1999;   99US-0137060.
PR   19-JAN-2000; 2000US-0176993.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Zoghbi HY,  Bellen H,  Birmingham N,  Hassan B,  Ben-Arie N;
XX
DR   WPI; 2001-032190/04.
DR   P-PSDB; AAB60357.
XX
PT   Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT   any of its homologs or orthologs, for the treatment of e.g. deafness,
PT   osteoarthritis and abnormal cell proliferation -
XX
PS   Disclosure; Page -; 142pp; English.
XX
CC   The invention relates to the use of atonal-associated nucleic acid or
CC   amino acid sequence, or any of its homologues or orthologues as
CC   therapeutic agents for the treatment of deafness, partial hearing loss,
CC   vestibular effects due to damage or loss of inner hair cells,
CC   osteoarthritis and abnormal cell proliferation. The invention also
CC   encompasses methods of screening for compounds which affect the
CC   expression of an atonal-associated nucleic acid sequence in an animal,
CC   and a transgenic animal in which an allele of a native atonal-associated
CC   gene is replaced by a heterologous nucleic acid sequence, thus
CC   inactivating the atonal-associated allele. The nucleic acids or proteins
CC   may be used in a method of treating an animal for hearing impairment,
CC   joint disease, balance disorders, abnormal cell proliferation, or other
CC   disease related to loss of a functional atonal-associated nucleic acid or
CC   protein. They may particularly be used to treat an animal with a
CC   deficiency in cerebellar granule neurons or their precursors, and may
CC   also be used in promoting mechanoreceptive cell growth and generating
CC   hair cells. The present sequence represents an atonal-associated nucleic
CC   acid sequence referred to in the invention.
CC   Note: The present sequence is not shown in the specification, but
CC   was obtained from GenBank.
XX
SQ   Sequence 790 BP; 91 A; 351 C; 283 G; 65 T; 0 other;

Alignment Scores:
Pred. No.:              7.19e-16        Length:         790
Score:                  388.50          Matches:        104
Percent Similarity:     54.31%          Conservative:   22
Best Local Similarity:  44.83%          Mismatches:     63
Query Match:            34.47%          Indels:         43
DB:                     22              Gaps:           10

US-09-595-947E-10 (1-214) x AAF27264 (1-790)

Qy          3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
              |||   |||          ||||||      |||   :::                   |||
Db         47 CCAGCGCCCCCCGACCGCCCGACGGCC---GTCGACCAGCGGCCGCGCCCCCCTCTCCCC 103
```

```
Qy         23 Arg-----------------------------------AlaSerGluAspGluVal 29
              |||                                   |||:::|||||||||:::
Db        104 CGCCCCCGCAGGATGCCGGTGAAGGCGGAGAGCCCGGCGCCCGCGGCGGAGGACGAACTG 163

Qy         30 ThrCysProThrSerAlaProProSerProThrArgThr---ProGlyAsnCysAlaGlu 48
                       |||   |||:::|||:::   :::   |||   :::       |||
Db        164 CTGCTGCTGCGCCTCGCCTCGCCCGCCCCCTCGGCCTCGCTGCCGTCCAGCGCCGGCGAG 223

Qy         49 AlaGluGluGlyGlyCysArgGlyAlaProArgLysLeuArg------AlaArgArgGly 66
                 :::|||           |||   ||||||:::|||:::      |||||||||
Db        224 GAGGACGAGGACGAGGAGGACGGGCGGCCGCGGCGGCTGCAGGAGGGCGCTCGGCGGGCG 283

Qy         67 GlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln---------------Arg 81
              ||||||   |||                 |||   :::   :::               :::
Db        284 GGGCGGCAGCGAGGGCCCCCGCGGGCGGCGCGCACGGCGGAGACGGCGCAGCGCATCAAG 343

Qy         82 ArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101
              ||||||||||||   |||||||||:::||||||||||||||||||:::|||||||:::
Db        344 CGGAGCCGGCGGCTGAAAGCCAACAACCGCGAGCGCAACCGCATGCACAACCTGAACGCG 403

Qy        102 AlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLys 121
              ||||||||||||||||   ||||||||||||||||:::|||||||||||||||||||
Db        404 GCGCTGGACGCGCTGCGCGACGTGCTGCCCACCTTCCCCGAGGACGCCAAGCTCACCAAG 463

Qy        122 IleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIle 141
              |||||||||||||||||||||||||||||||||||||||||:::|||||||:::
Db        464 ATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTCACCGAGACGCTGCGCCTG 523

Qy        142 AlaAspHisSer---LeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              |||      :::   |||                 ||||||              |||
Db        524 GCCGGGGCCGCCCGCCTGGGGGGCGCCGCCGACGCCGCGCCCGGGGCGGCCGCCGAGGGC 583

Qy        161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
              ||||||      |||         |||                          :::|||
Db        584 AGCCCC---TCGCCCGCCTCGTCGTGG-------------------------AGCGGC 613

Qy        181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
                    |||||||||   |||                              |||:::
Db        614 GGCGCCAGCCCCCGCGCCCTCC-------------------------GCCTCGCCCTAC 646

Qy        201 AlaCys---LeuSerProGlySerLeuAlaPheSer 211
              ||||||   ||||||||||||||   |||   |||
Db        647 GCCTGCACTTTATCGCCCGGCAGCCCCGCCGGCTCC 682
```

Search completed: February 2, 2004, 20:02:31
Job time : 283 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 2, 2004, 18:55:26 ; Search time 355 Seconds
(without alignments)
2197.000 Million cell updates/sec

Title: US-09-595-947E-10
Perfect score: 1127
Sequence: 1 MTPQPSGAPTVQVTRETERS.....LGATSSACLSPGSLAFSDFL 214

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09595947/runat_02022004_154934_8079/app_query.fasta_1.391
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09595947_@CGN_1_1_387_@runat_02022004_154934_8079
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1105	98.0	5340	9	US-09-817-360-1	Sequence 1, Appli
2	819	72.7	645	14	US-10-004-717-4	Sequence 4, Appli
3	819	72.7	861	14	US-10-004-717-24	Sequence 24, Appl
4	819	72.7	1861	9	US-09-817-360-3	Sequence 3, Appli
5	388.5	34.5	790	14	US-10-004-717-20	Sequence 20, Appl
6	385.5	34.2	1074	14	US-10-004-717-18	Sequence 18, Appl
7	375.5	33.3	1385	14	US-10-004-717-30	Sequence 30, Appl
8	370.5	32.9	1412	14	US-10-004-717-6	Sequence 6, Appli
9	370.5	32.9	1412	14	US-10-004-717-37	Sequence 37, Appl
10	370	32.8	1527	8	US-08-722-570-12	Sequence 12, Appl
11	359.5	31.9	738	8	US-08-722-570-13	Sequence 13, Appl
12	322.5	28.6	1312	8	US-08-722-570-14	Sequence 14, Appl
13	304	27.0	1277	8	US-08-722-570-15	Sequence 15, Appl
14	238.5	21.2	1550	14	US-10-004-717-43	Sequence 43, Appl
15	238.5	21.2	1957	14	US-10-004-717-8	Sequence 8, Appli
16	215	19.1	1099	13	US-10-413-358-27	Sequence 27, Appl
17	215	19.1	1211	13	US-10-413-358-26	Sequence 26, Appl
18	207.5	18.4	993	14	US-10-004-717-47	Sequence 47, Appl
19	207.5	18.4	3261	14	US-10-004-717-12	Sequence 12, Appl
20	207.5	18.4	3541	14	US-10-004-717-32	Sequence 32, Appl
c 21	206	18.3	1021	12	US-10-321-039-71	Sequence 71, Appl
22	205.5	18.2	1056	14	US-10-004-717-10	Sequence 10, Appl
23	205.5	18.2	1393	14	US-10-004-717-45	Sequence 45, Appl
24	198.5	17.6	675	14	US-10-004-717-15	Sequence 15, Appl
25	198.5	17.6	2196	12	US-10-108-260A-1239	Sequence 1239, Ap
26	197.5	17.5	485	14	US-10-004-717-59	Sequence 59, Appl
27	194	17.2	748	13	US-10-029-386-25014	Sequence 25014, A
28	192.5	17.1	1065	14	US-10-004-717-1	Sequence 1, Appli
29	192.5	17.1	1572	14	US-10-004-717-57	Sequence 57, Appl
30	191.5	17.0	2315	12	US-10-136-728-77	Sequence 77, Appl
31	190	16.9	501	14	US-10-004-717-13	Sequence 13, Appl
32	187	16.6	948	14	US-10-004-717-41	Sequence 41, Appl
33	187	16.6	2993	15	US-10-125-237-56	Sequence 56, Appl
34	187	16.6	2993	15	US-10-105-891-56	Sequence 56, Appl
35	185.5	16.5	849	13	US-10-226-872-8	Sequence 8, Appli
36	183.5	16.3	907	14	US-10-004-717-65	Sequence 65, Appl
37	183.5	16.3	932	12	US-10-295-027-1190	Sequence 1190, Ap
38	183.5	16.3	1791	13	US-10-226-872-1	Sequence 1, Appli
c 39	183.5	16.3	1791	13	US-10-226-872-6	Sequence 6, Appli
40	183.5	16.3	1830	13	US-10-226-872-4	Sequence 4, Appli
c 41	183.5	16.3	17290	11	US-09-999-121-7	Sequence 7, Appli
c 42	183.5	16.3	25760	11	US-09-999-121-13	Sequence 13, Appl
43	180	16.0	1034	13	US-10-029-386-22762	Sequence 22762, A
44	179	15.9	849	13	US-10-226-872-9	Sequence 9, Appli
45	177.5	15.7	938	14	US-10-004-717-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1
US-09-817-360-1
; Sequence 1, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5340
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-817-360-1

```
Alignment Scores:
Pred. No.:              6.88e-92        Length:         5340
Score:                  1105.00         Matches:        211
Percent Similarity:     99.07%          Conservative:   1
Best Local Similarity:  98.60%          Mismatches:     2
Query Match:            98.05%          Indels:         0
DB:                     9               Gaps:           0

US-09-595-947E-10 (1-214) x US-09-817-360-1 (1-5340)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3022 ATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGAGCGGTCC 3081

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3082 TTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCCAGCCCCACT 3141

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3142 CGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCCCCGAGGAAG 3201

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3202 CTCCGGGCACGGCGCGGGGGACGCAGCCGGCCTAAGAGCGAGTTGGCACTGAGCAAGCAG 3261

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       3262 CGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACAACCTCAAC 3321

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3322 TCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCGAAGCTCACC 3381

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3382 AAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAACGCTGCGC 3441

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3442 ATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGAGCTGGGC 3501

Qy        161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
              ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db       3502 AGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCAGGCTGGC 3561

Qy        181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3562 AGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCACCTCTTCC 3621

Qy        201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
              |||||||||||||||||||||||||||||||||||||||||||
Db       3622 GCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTG 3663

RESULT 2
US-10-004-717-4
; Sequence 4, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 645
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-4

Alignment Scores:
Pred. No.:              1.4e-66         Length:         645
Score:                  819.00          Matches:        163
Percent Similarity:     82.33%          Conservative:   14
Best Local Similarity:  75.81%          Mismatches:     36
Query Match:            72.67%          Indels:         2
DB:                     14              Gaps:           2

US-09-595-947E-10 (1-214) x US-10-004-717-4 (1-645)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||  |||  |||      |||  |||:::||||||:::  ||||||::::::
Db          1 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 60

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||  ||||||:::  ||||||              |||  ||||||||||||||||
Db         61 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 120

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                  |||   :::|||:::|||||||||   |||   |||||||||       ||||||
Db        121 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 180

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db        181 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 240

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        241 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 300

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 360

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 420

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||  |||      |||||||||  |||  ||||||||||||||||
Db        421 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 477

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::||||||||||||||||||||
Db        478 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 537

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   ||||||||||||||   |||||||||          :::|||
Db        538 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 597

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                 :::  |||  ||||||:::|||  ||||||||||||||
Db        598 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 642

RESULT 3
US-10-004-717-24
; Sequence 24, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
```

```
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 861
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-24

Alignment Scores:
Pred. No.:              1.9e-66       Length:          861
Score:                  819.00        Matches:         163
Percent Similarity:     82.33%        Conservative:    14
Best Local Similarity:  75.81%        Mismatches:      36
Query Match:            72.67%        Indels:          2
DB:                     14            Gaps:            2

US-09-595-947E-10 (1-214) x US-10-004-717-24 (1-861)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::|||||||:::   ||||||::::::
Db        160 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 219

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||              |||   ||||||||||||||
Db        220 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 279

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                  |||   :::|||:::|||||||||   |||   |||||||||       ||||||
Db        280 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 339

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        340 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 399

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        400 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 459

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 519

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 579

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||       |||||||||   |||   ||||||||||||||
Db        580 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 636

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::||||||||||||||||||||
Db        637 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 696

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   ||||||||||||||   |||||||||          :::|||
Db        697 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 756

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                 :::   |||   ||||||:::|||   ||||||||||||||
Db        757 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 801

RESULT 4
US-09-817-360-3
; Sequence 3, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
;  APPLICANT: German, Michael S.
;  APPLICANT:  Lin, Joseph
;  TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
;  TITLE OF INVENTION:  AND DELIVERY OF INSULIN
;  FILE REFERENCE: UCSF-129CIP
;  CURRENT APPLICATION NUMBER: US/09/817,360
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: 09/535,145
;  PRIOR FILING DATE: 2000-03-24
;  PRIOR APPLICATION NUMBER: 60/128,180
;  PRIOR FILING DATE: 1999-04-06
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1861
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-817-360-3

Alignment Scores:
Pred. No.:              4.33e-66      Length:          1861
Score:                  819.00        Matches:         163
Percent Similarity:     82.33%        Conservative:    14
Best Local Similarity:  75.81%        Mismatches:      36
Query Match:            72.67%        Indels:          2
DB:                     9             Gaps:            2

US-09-595-947E-10 (1-214) x US-09-817-360-3 (1-1861)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::|||||||:::   ||||||::::::
Db       1093 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 1152

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||              |||   ||||||||||||||
Db       1153 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 1212

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                  |||   :::|||:::|||||||||   |||   |||||||||       ||||||
Db       1213 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 1272

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db       1273 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 1332

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       1333 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 1392

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1393 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 1452

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1453 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 1512

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||   |||       |||||||||   |||   ||||||||||||||
Db       1513 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 1569

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   ||||||      ||||||||||||||:::||||||||||||||||||||
Db       1570 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 1629

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::|||||||||   ||||||||||||||   |||||||||          :::|||
Db       1630 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 1689

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                 :::   |||   ||||||:::|||   ||||||||||||||
```

```
Db       1690 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 1734

RESULT 5
US-10-004-717-20
; Sequence 20, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
;   LENGTH: 790
;   TYPE: DNA
;   ORGANISM: chicken
US-10-004-717-20

Alignment Scores:
Pred. No.:              6.44e-27      Length:          790
Score:                  388.50        Matches:         104
Percent Similarity:     54.31%        Conservative:    22
Best Local Similarity:  44.83%        Mismatches:      63
Query Match:            34.47%        Indels:          43
DB:                     14            Gaps:            10

US-09-595-947E-10 (1-214) x US-10-004-717-20 (1-790)

Qy           3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
               |||   |||          ||||||      |||   :::                    |||
Db          47 CCAGCGCCCCCCGACCGCCCGACGGCC---GTCGACCAGCGGCCGCGCCCCCCTCTCCCC 103

Qy          23 Arg-------------------------------------AlaSerGluAspGluVal 29
               |||                                     |||:::|||||||||:::
Db         104 CGCCCCCGCAGGATGCCGGTGAAGGCGGAGAGCCCGGCGCCCGCGGCGGAGGACGAACTG 163

Qy          30 ThrCysProThrSerAlaProProSerProThrArgThr---ProGlyAsnCysAlaGlu 48
                     |||   |||:::|||:::   :::   |||   :::      |||
Db         164 CTGCTGCTGCGCCTCGCCTCGCCCGCCCCCTCGGCCTCGCTGCCGTCCAGCGCCGGCGAG 223

Qy          49 AlaGluGluGlyGlyCysArgGlyAlaProArgLysLeuArg------AlaArgArgGly 66
                  :::|||          |||   ||||||:::|||:::      |||||||||
Db         224 GAGGACGAGGACGAGGAGGACGGGCGGCCGCGGCGGCTGCAGGAGGGCGCTCGGCGGGCG 283

Qy          67 GlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln--------------Arg 81
               ||||||   |||                |||   :::   :::               :::
Db         284 GGGCGGCAGCGAGGGCCCCCGCGGGCGGCGCGCACGGCGGAGACGGCGCAGCGCATCAAG 343

Qy          82 ArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101
               ||||||||||||   |||||||||:::||||||||||||||||||:::||||||:::
Db         344 CGGAGCCGGCGGCTGAAAGCCAACAACCGCGAGCGCAACCGCATGCACAACCTGAACGCG 403

Qy         102 AlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLys 121
               ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db         404 GCGCTGGACGCGCTGCGCGACGTGCTGCCCACCTTCCCCGAGGACGCCAAGCTCACCAAG 463

Qy         122 IleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIle 141
               ||||||||||||||||||||||||||||||||||||||||||:::|||||||||:::
Db         464 ATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTCACCGAGACGCTGCGCCTG 523

Qy         142 AlaAspHisSer---LeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
               |||      :::   |||                  ||||||                |||

Db         524 GCCGGGGCCGCCCGCCTGGGGGGCGCCGCCGACGCCGCGCCCGGGGCGGCCGCCGAGGGC 583

Qy         161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
               ||||||      |||          |||                        :::|||
Db         584 AGCCCC---TCGCCCGCCTCGTCGTGG--------------------------AGCGGC 613

Qy         181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
                     |||||||||   |||                                |||:::
Db         614 GGCGCCAGCCCCGCGCCCTCC--------------------------GCCTCGCCCTAC 646

Qy         201 AlaCys---LeuSerProGlySerLeuAlaPheSer 211
               ||||||   ||||||||||||||||   |||   |||
Db         647 GCCTGCACTTTATCGCCCGGCAGCCCCGCCGGCTCC 682

RESULT 6
US-10-004-717-18
; Sequence 18, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 1074
;   TYPE: DNA
;   ORGANISM: chicken
US-10-004-717-18

Alignment Scores:
Pred. No.:              1.69e-26      Length:          1074
Score:                  385.50        Matches:         104
Percent Similarity:     51.29%        Conservative:    15
Best Local Similarity:  44.83%        Mismatches:      60
Query Match:            34.21%        Indels:          53
DB:                     14            Gaps:            10

US-09-595-947E-10 (1-214) x US-10-004-717-18 (1-1074)

Qy          21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerPro--- 39
               ||||||      |||                |||          |||||||:::|||
Db         228 TTTCCCTCCCTTTCCCCCTCC-----------CCTCAGCACCTTCCCCCCGCTCCCCCT 275

Qy          40 ----------------------------ThrArgThrProGlyAsnCysAlaGluAla 49
                                           |||   |||         |||   :::
Db         276 GCCGCCCACTGACGGCTCCGGCTTCTCTCCCGCAGGATGCCCGCGGAGGCGGCGAGCAGC 335

Qy          50 GluGluGlyGlyCysArg---------GlyAlaProArgLysLeuArgAlaArgArgGly 66
                     ||||||              ||||||||||||:::   |||   ||||||
Db         336 ------GGCGGCGTTTCGGAGCCGCCCGGAGCTCCGCGGGAG---CGGCGGAGGAGACGC 386

Qy          67 GlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArgLys 86
               ||||||:::|||   ::::::|||      |||          :::||||||||||
Db         387 GGCCGTGCGCGGGCGCGGACCGAGGCTTTGCTGCACACCCTCAAACGGAGCCGCCGGGTG 446

Qy          87 LysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeu 106
               ||||||||||||||||||||||||||||||   ||||||:::|||||||   |||
Db         447 AAAGCCAACGACCGGGAGCGGAACCGCATGCACCACCTCAACGCCGCGCTGGATGAGCTC 506

Qy         107 ArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeuArg 126
               |||   ||||||||||||||||||||||||   ||||||||||||||||||||||||||
```

```
Db        507 CGCAGCGTCCTGCCGACCTTCCCCGACGACACCAAACTCACCAAAATCGAAACCCTGCGC 566

Qy        127 PheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu 146
              ||||||:::|||||||||||||||||||:::::||||||||:::|||:::      |||
Db        567 TTCGCTTACAACTACATCTGGGCCCTCTCCGAGACCCTTCGTTTGGCCGAGCAGTGCCTC 626

Qy        147 ------------TyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySerPro 162
                          ||||||||||||                ||||||
Db        627 CCTCCTCCCCCCGCCTTCCGCGGGCCCCCCGCGCCC---------------CCCAGCCCC 671

Qy        163 GlyGlyProProGlyAspTrpGlySerLeuTyrSerPro------------------Val 176
              |||         |||  |||   |||      ||||||
Db        672 GGCAGCGACGCCGGTTCGTGGCTGTCCAGCGGTTCCCCGGCCGCCCCCTCGCTCTGCGCC 731

Qy        177 SerGlnAlaGlySerLeuSerProAlaAlaSer--------------------------- 187
              |||   :::|||      |||||||||  |||
Db        732 TCCGCCTCCGGCCCCAGCAGCCCGGCCACCTCCGAGGACTGCGGCTACGTCCCCTCGGAC 791

Qy        188 ---LeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAla--------------- 201
                 |||             |||         |||          |||
Db        792 GCCCTGCGGGCCTTCCGCGGGCTGCCCCCCGCCGCCCCGGGCGCTCCCTGCCGCTAGCCC 851

Qy        202 ------CysLeuSerProGlySerLeuAlaPheSer 211
                   |||:::||||||          ||||||
Db        852 TGCCCGTGCGTGTCTCCGTCCCCCCCCACCTTCTCC 887

RESULT 7
US-10-004-717-30
; Sequence 30, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
;   LENGTH: 1385
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-30

Alignment Scores:
Pred. No.:              1.84e-25      Length:        1385
Score:                  375.50        Matches:       104
Percent Similarity:     50.00%        Conservative:  19
Best Local Similarity:  42.28%        Mismatches:    55
Query Match:            33.32%        Indels:        69
DB:                     14            Gaps:          10

US-09-595-947E-10 (1-214) x US-10-004-717-30 (1-1385)

Qy         26 GluAspGluVal-----------------ThrCysProThrSerAlaProProSerPro 39
              |||:::||||||                |||   ||||||     :::|||
Db        418 GAAGAGGAGGTACTGATGCTGCTGGGCTCGGCTTCCCCGGCCTCGGCGACCCTGACCCCG 477

Qy         40 ----------------------------------ThrArgThrProGlyAsnCys--- 46
                                                |||   ||||||:::
Db        478 ATGTCCTCCAGCGCGGACGAGGAGGAGGACGAGGAGCTGCGCCGGCCGGGCTCCGCGCGT 537

Qy         47 ------------AlaGluAlaGluGlu------------------------------Gly 52
                         ||||||||||||:::                              |||

Db        538 GGGCAGCGTGGAGCGGAAGCCGAGCAGGGGGTGCAGGGCAGTCCGGCGTCGGGTGCCGGG 597

Qy         53 GlyCysArgGlyAlaProArgLysLeuArgAla-----------------ArgArgGly 66
              |||||||||      |||   :::|||                      ||||||
Db        598 GGTTGCCGG------CCAGGGCGGCTGCTGGGCCTGATGCACGAGTGCAAGCGTCGCCCG 651

Qy         67 GlyArgSerArg---------------ProLysSerGluLeuAlaLeuSerLysGlnArg 81
              |||||||||                |||:::          :::   :::   :::
Db        652 TCGCGCTCACGGGCCGTCTCCCGAGGTGCCAAGACGGCGGAGACGGTGCAGCGCATCAAG 711

Qy         82 ArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101
              ::::::||||||  |||||||||:::||||||||||||||||||||:::||||||:::
Db        712 AAGACCCGCAGGCTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCC 771

Qy        102 AlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLys 121
              ||||||||||||||||||  ||||||||||||||||||:::||||||||||||||||||
Db        772 GCGCTGGACGCGCTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAG 831

Qy        122 IleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIle 141
              ||||||||||||||||||||||||||||||||||||||||||||:::|||||||||:::
Db        832 ATCGAGACGCTGCGCTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTG 891

Qy        142 AlaAspHisSerLeuTyrAla-----------------------------------Leu 149
              |||||||||       |||                                     |||
Db        892 GCGGACCACTGCGCCGGCGCCGGTGGCCTCCAGGGGGCGCTCTTCACGGAGGCGGTGCTC 951

Qy        150 GluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTrp 169
                 |||         |||      |||           ||||||      |||
Db        952 CTGAGCCCGGGAGCTGCGCTCGGCGCCAGCGGGGACAGCCCTTCTCCACCTTCCTCCTGG 1011

Qy        170 GlySerLeuTyrSerProValSerGlnAlaGlySerLeuSerPro--------------- 184
                          ||||||   |||   :::   |||   ||||||
Db       1012 AGCTGCACCAACAGCCCGGCGTCATCCTCCAACTCCACGTCCCCATACAGCTGCACTTTA 1071

Qy        185 ---AlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAla------ 201
                    ||||||      |||   ||||||||   |||   :::|||
Db       1072 TCGCCCGCTAGCCC-CGGGTCAGACGTGGACTACTGGCAGCCCCCACCTCCGGAGAAGCA 1130

Qy        202 ------CysLeuSerPro 205
                   |||   ||||||
Db       1131 TCGTTATGCGCCTCACCT 1148

RESULT 8
US-10-004-717-6
; Sequence 6, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1412
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-6

Alignment Scores:
Pred. No.:              5.4e-25       Length:        1412
Score:                  370.50        Matches:       106
Percent Similarity:     48.47%        Conservative:  21
```

```
Best Local Similarity:  40.46%        Mismatches:    61
Query Match:            32.87%        Indels:        75
DB:                     14            Gaps:          11

US-09-595-947E-10 (1-214) x US-10-004-717-6 (1-1412)

Qy         16 GluThrGluArgSerPheProArgAla------------------SerGluAspGluVal 29
              |||       |||   |||   ::::::                  |||:::||||||
Db         73 GAGCCGCGTAGGATGTTCGTCAAATCTGAGACTCTGGAGTTGAAGGAGGAAGAGGAGGTA 132

Qy         30 ------------------ThrCysProThrSerAlaProProSerPro------------ 39
                                  |||   ||||||      :::|||
Db        133 CTGATGCTGCTGGGCTCGGCTTCCCCGGCCTCGGCGACCCTGACCCCGATGTCCTCCAGC 192

Qy         40 ------------------------ThrArgThrProGlyAsnCys--------------- 46
                                     |||   ||||||:::
Db        193 GCGGACGAGGAGGAGGACGAGGAGCTGCGCCGGCCGGGCTCCGCGCGTGGGCAGCGTGGA 252

Qy         47 AlaGluAlaGluGlu-----------------------------GlyGlyCysArgGly 56
              |||||||||   :::                            ||||||||||
Db        253 GCGGAAGCCGGGCAGGGGGTGCAGGGCAGTCCGGCGTCGGGTGCCGGGGGTTGCCGG--- 309

Qy         57 AlaProArgLysLeuArgAla------------------ArgArgGlyGlyArgSerArg 70
                 |||   :::|||                        ||||||      |||||||||
Db        310 ---CCAGGGCGGCTGCTGGGCCTGATGCACGAGTGCAAGCGTCGCCCGTCGCGCTCACGG 366

Qy         71 ---------------ProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArg 85
                           |||:::      :::  :::  :::::::::||||||
Db        367 GCCGTCTCCCGAGGTGCCAAGACGGCGGAGACGGTGCAGCGCATCAAGAAGACCCGCAGG 426

Qy         86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
              |||||||||:::||||||||||||||||||||:::|||||||:::||||||||||||
Db        427 CTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCGCTGGACGCG 486

Qy        106 LeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeu 125
              ||||||   ||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        487 CTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATCGAGACGCTG 546

Qy        126 ArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSer 145
              |||||||||||||||||||||||||||||||:::|||||||||:::|||||||||||
Db        547 CGCTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCGGACCACTGC 606

Qy        146 LeuTyrAla-----------------------------------LeuGluProProAla 153
                 |||                                     |||      |||
Db        607 GCCGGCGCCGGTGGCCTCCAGGGGGCGCTCTTCACGGAGGCGGTGCTCCTGAGCCCGGGA 666

Qy        154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTrpGlySerLeuTyr 173
                 |||      |||               ||||||      |||
Db        667 GCTGCGCTCGGCGCCAGCGGGGACAGCCCTTCTCCACCTTCCTCCTGGAGCTGCACCAAC 726

Qy        174 SerProValSerGlnAlaGlySerLeuSerPro------------------AlaAlaSer 187
              ||||||   |||  :::   |||   ||||||                  ||||||
Db        727 AGCCCGGCGTCATCCTCCAACTCCACGTCCCCATACAGCTGCACTTTATCGCCCGCTAGC 786

Qy        188 LeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAla------------CysLeu 203
                 |||   |||||||||   |||   :::|||                    |||
Db        787 CC-CGGGTCAGACGTGGACTACTGGCAGCCCCCACCTCCGGAGAAGCATCGTTATGCGCC 845

Qy        204 SerPro 205
              ||||||
Db        846 TCACCT 851

RESULT 9
US-10-004-717-37
; Sequence 37, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
;   LENGTH: 1412
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-37

Alignment Scores:
Pred. No.:              5.4e-25       Length:        1412
Score:                  370.50        Matches:       106
Percent Similarity:     48.47%        Conservative:  21
Best Local Similarity:  40.46%        Mismatches:    61
Query Match:            32.87%        Indels:        75
DB:                     14            Gaps:          11

US-09-595-947E-10 (1-214) x US-10-004-717-37 (1-1412)

Qy         16 GluThrGluArgSerPheProArgAla------------------SerGluAspGluVal 29
              |||       |||   |||   ::::::                  |||:::||||||
Db         73 GAGCCGCGTAGGATGTTCGTCAAATCTGAGACTCTGGAGTTGAAGGAGGAAGAGGAGGTA 132

Qy         30 ------------------ThrCysProThrSerAlaProProSerPro------------ 39
                                  |||   ||||||      :::|||
Db        133 CTGATGCTGCTGGGCTCGGCTTCCCCGGCCTCGGCGACCCTGACCCCGATGTCCTCCAGC 192

Qy         40 ------------------------ThrArgThrProGlyAsnCys--------------- 46
                                     |||   ||||||:::
Db        193 GCGGACGAGGAGGAGGACGAGGAGCTGCGCCGGCCGGGCTCCGCGCGTGGGCAGCGTGGA 252

Qy         47 AlaGluAlaGluGlu-----------------------------GlyGlyCysArgGly 56
              |||||||||   :::                            ||||||||||
Db        253 GCGGAAGCCGGGCAGGGGGTGCAGGGCAGTCCGGCGTCGGGTGCCGGGGGTTGCCGG--- 309

Qy         57 AlaProArgLysLeuArgAla------------------ArgArgGlyGlyArgSerArg 70
                 |||   :::|||                        ||||||      |||||||||
Db        310 ---CCAGGGCGGCTGCTGGGCCTGATGCACGAGTGCAAGCGTCGCCCGTCGCGCTCACGG 366

Qy         71 ---------------ProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArg 85
                           |||:::      :::  :::  :::::::::||||||
Db        367 GCCGTCTCCCGAGGTGCCAAGACGGCGGAGACGGTGCAGCGCATCAAGAAGACCCGCAGG 426

Qy         86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
              |||||||||:::||||||||||||||||||||:::|||||||:::||||||||||||
Db        427 CTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCGCTGGACGCG 486

Qy        106 LeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeu 125
              ||||||   ||||||||||||||||||:::|||||||||||||||||||||||||||||
Db        487 CTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATCGAGACGCTG 546

Qy        126 ArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSer 145
              |||||||||||||||||||||||||||||||:::|||||||||:::|||||||||||
Db        547 CGCTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCGGACCACTGC 606

Qy        146 LeuTyrAla-----------------------------------LeuGluProProAla 153
                 |||                                     |||      |||
Db        607 GCCGGCGCCGGTGGCCTCCAGGGGGCGCTCTTCACGGAGGCGGTGCTCCTGAGCCCGGGA 666

Qy        154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTrpGlySerLeuTyr 173
                 |||      |||               ||||||      |||
Db        667 GCTGCGCTCGGCGCCAGCGGGGACAGCCCTTCTCCACCTTCCTCCTGGAGCTGCACCAAC 726

Qy        174 SerProValSerGlnAlaGlySerLeuSerPro------------------AlaAlaSer 187
              ||||||   |||  :::   |||   ||||||                  ||||||
Db        727 AGCCCGGCGTCATCCTCCAACTCCACGTCCCCATACAGCTGCACTTTATCGCCCGCTAGC 786
```

```
Qy        188 LeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAla-----------CysLeu 203
                  |||   |||||||||   |||   :::|||               |||
Db        787 CC-CGGGTCAGACGTGGACTACTGGCAGCCCCCACCTCCGGAGAAGCATCGTTATGCGCC 845

Qy        204 SerPro 205
                  ||||||
Db        846 TCACCT 851

RESULT 10
US-08-722-570-12
; Sequence 12, Application US/08722570
; Publication No. US20030044887A1
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  5365
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1527 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-12

Alignment Scores:
Pred. No.:              6.53e-25        Length:         1527
Score:                  370.00          Matches:        99
Percent Similarity:     52.97%          Conservative:   17
Best Local Similarity:  45.21%          Mismatches:     63
Query Match:            32.83%          Indels:         40
DB:                     8               Gaps:           8

US-09-595-947E-10 (1-214) x US-08-722-570-12 (1-1527)

Qy          3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
              |||      :::      |||      |||:::      |||   |||:::   |||
Db        293 CCAGCAGCAACAGCGGGAGCGACCTGTCCAGTTTCCTCACCGACGAGGAGGACTGTGCCA 352

Qy         23 ArgAlaSerGlu------AspGluValThrCysProThr-----------SerAlaPro 36
                 :::|||                |||||                  :::   |||
Db        353 GGCTCCAGCCCCTAGCTTCCACCTCAGGGCTGTCCGTGCCAGCCCGCAGGAGCGCGCCCA 412

Qy         37 ProSerPro---ThrArgThr--ProGlyAsnCysAlaGluAlaGluGluGlyGlyCysA 55
              |||||||||      ||||||  ||||||         |||   :::|||
Db        413 CCCTCTCCGGGGCATCGAACGTTCCCGGTGGCCAGGACGAAGAGCAGGAG---------- 462

Qy         55 rgGlyAlaProArgLysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluL 75
                                  |||   ||||||   ||||||:::|||   :::||||||
Db        463 -------------------CGGCGGCGACGGCGAGGTCGCGCGCGGGTGCGGTCCGAGG 502

Qy         75 euAlaLeuSerLysGlnArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnA 95
                |||          ||||||||||||||   ||||||||||||||||||||||||
Db        503 CGCTGCTGCACTCGCTGCGGAGGAGCCGTCGCGTCAAGGCCAACGATCGCGAGCGCAACC 562

Qy         95 rgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProA 115
              ||||||||:::|||||||:::||||||||||||||||||   |||||||||:::|||||||
Db        563 GTATGCATAACCTCAACGCTGCGCTGGACGCTCTGCGCAGCGTGCTGCCCTCGTTCCCCG 622

Qy        115 spAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaL 135
              |||||   ||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db        623 ACGACACCAAGCTCACCAAGATTGAGACGCTGCGCTTCGCCTACAACTACATCTGGGCCC 682

Qy        135 euThrGlnThrLeuArgIleAlaAspHisSerLeu--------------------TyrA 148
              ||   :::|||||||||:::||||||      |||
Db        683 TGGCTGAGACACTGCGCCTGGCAGATCAAGGGCTCCCGGGGGGCGGTGCCCGGGAGCGCC 742

Qy        148 laLeuGluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyProProGlyA 168
               |||   |||         |||      |||      |||      |||      |
Db        743 TCCTGCCTCCGCAGTGTGTCCCCTGC------CTGCCCGGTCCCCCGAGCCCGGCCAGCG 796

Qy        168 sp---------TrpGlySer-------------------------LeuTyrSerProV 176
              ||         |||||||||                      :::   ||||||:
Db        797 ATACAGAGTCCTGGGGCTCCGGGGCCGCTGCCTCCCCCTGCGCTACTGTGGCGTCACCAC 856

Qy        176 alSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgProGly 193
              ::|||         |||   |||   :::                 ||||||
Db        857 TCTCTGACCCCAGTAGTCCCTCGGCTTCAGAAGACTTCACCTATGGCCCGGGT 909

RESULT 11
US-08-722-570-13
; Sequence 13, Application US/08722570
; Publication No. US20030044887A1
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  5365
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-13
```

```
Alignment Scores:
Pred. No.:              2.77e-24        Length:         738
Score:                  359.50          Matches:        97
Percent Similarity:     51.83%          Conservative:   16
Best Local Similarity:  44.50%          Mismatches:     66
Query Match:            31.90%          Indels:         40
DB:                     8               Gaps:           8

US-09-595-947E-10 (1-214) x US-08-722-570-13 (1-738)

Qy          3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
              |||      :::   |||   |||         :::      |||   |||:::   |||
Db         44 CCAGCAGCAACAGCAGCAGCGACCTGTCCAGCTTCCTCACCGACGAGGAGGACTGTGCCA 103

Qy         23 ArgAlaSer------GluAspGluValThrCysProThr-----------SerAlaPro 36
                  |||              |||||                    :::   |||
Db        104 GGCTACAGCCCCTAGCCTCCACCTCGGGGCTGTCCGTGCCAGCCCGGAGGAGCGCTCCCG 163

Qy         37 ProSerProThr------ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCys 54
              |||||||||              ||||||         |||   :::|||
Db        164 CCCTCTCCGGG-GCATCGAATGTTCCCGGTGCCCAGGACGAAGAGCAGGAA--------- 213

Qy         55 ArgGlyAlaProArgLysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGlu 74
                          |||   ||||||   ||||||:::|||   :::||||||
Db        214 --------------------CGGCGGAGGCGGCGAGGTCGCGCTCGGGTGCGGTCCGAG 252

Qy         75 LeuAlaLeuSerLysGlnArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsn 94
              |||          ||||||||||||||   ||||||||||||||||||||||
Db        253 GCTCTGCTGCACTCCCTGCGGAGGAGTCGTCGCGTCAAAGCCAACGATCGCGAGCGCAAC 312

Qy         95 ArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPhePro 114
              |||||||||:::||||||:::||||||||||||||||||   |||||||||:::||||||
Db        313 CGCATGCACAACCTCAACGCTGCGCTGGACGCCTTGCGCAGCGTGCTGCCCTCGTTCCCC 372

Qy        115 AspAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAla 134
              ||||||   ||||||||||||||||||||||||||||||:::||||||||||||||||
Db        373 GACGACACCAAGCTCACCAAGATTGAGACGCTGCGCTTCGCCTACAACTACATCTGGGCC 432

Qy        135 LeuThrGlnThrLeuArgIleAlaAspHisSerLeu--------------------Tyr 147
              |||   :::|||||||||:::||||||    |||
Db        433 CTGGCTGAGACACTGCGCCTGGCAGATCAAGGGCTCCCCGGGGGCAGTGCCCGGGAGCGC 492

Qy        148 AlaLeuGluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyProProGly 167
                 |||   |||           |||       |||       |||       |||
Db        493 CTCCTGCCTCCGCAGTGTGTCCCCTGT------CTGCCCGGGCCCCCGAGCCCGGCCAGC 546

Qy        168 Asp---------TrpGlySer--------------------------LeuTyrSerPro 175
              |||         |||||||||                        :::   ||||||
Db        547 GACACTGAGTCCTGGGGTTCCGGGGCCGCTGCCTCCCCCTGCGCCACTGTGGCATCACCA 606

Qy        176 ValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgProGly 193
              :::|||     |||   |||   :::                  ||||||
Db        607 CTCTCTGACCCCAGTAGTCCCTCGGCTTCAGAAGACTTCACCTATGGCCCGGGC 660

RESULT 12
US-08-722-570-14
; Sequence 14, Application US/08722570
; Publication No. US20030044887A1
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  5365
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1312 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-14

Alignment Scores:
Pred. No.:              1.29e-20        Length:         1312
Score:                  322.50          Matches:        86
Percent Similarity:     54.46%          Conservative:   30
Best Local Similarity:  40.38%          Mismatches:     83
Query Match:            28.62%          Indels:         14
DB:                     8               Gaps:           4

US-09-595-947E-10 (1-214) x US-08-722-570-14 (1-1312)

Qy          3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
              |||:::      |||   |||   :::   :::           :::      ||||||
Db        380 CCGCGATGAAGAGGAGGACCTGACCTCTGCCTCCCCCTGCTCCGTGACCTCCTCTTTCCG 439

Qy         23 ArgAlaSerGluAspGluValThrCysProThrSerAlaProPro-SerProThrArgTh 42
                        :::   :::|||              |||   ||| |||||||||   ::
Db        440 TTCCCCGGCGACGCAGACGTGCAGCTCGGACGATGAGCAGCTCCTGAGTCCCACC---AG 496

Qy         42 rProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLysLeuAr 62
              :||||||              :::      |||      :::||||||      ||
Db        497 CCCGGGA--------------CAGCACCAGGGGGAAGAGAACAGCCCGCGATGCAGGAG 541

Qy         62 gAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgAr 82
              |:::|||      ||||||::::::  ||||||          :::   |||   :::::
Db        542 GAGCCGA------GGCCGCGCTCAGGGCAAGAGCGGAGAAACTGTGTTAAAGATCAAGAA 595

Qy         82 gSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAl 102
              ::::||||||   ||||||||||:::||||||||||||||||||||:::||||||||||
Db        596 GACCCGGCGCGTTAAAGCTAACAACCGGGAAAGGAATCGCATGCACAACCTGAACTCTGC 655

Qy        102 aLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIl 122
              |||||||:::||||||   |||||||||:::   |||:::||||||||||||||||||||
Db        656 GCTTGATTCCCTCAGGGAAGTGTTGCCCTCTTTACCTGAAGATGCCAAACTCACCAAGAT 715

Qy        122 eGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAl 142
              ||||||||||||||||||:::||||||||||||||||||:::::|||||||||:::
Db        716 AGAGACCTTGCGCTTTGCCTACAACTACATCTGGGCTCTTAGCGAAACTTTGCGCCTTGG 775

Qy        142 aAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySerPr 162
              |||                :::          ||||||                  |||
Db        776 CGACCCAGTGCACCGATCTGCTTCCACCCCAGCAGCAGCCATATTGGTGCAGGACTCCTC 835

Qy        162 oGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGlySerLe 182
                                  |||          ||||||   |||          |||
Db        836 TTCATCCCAGAGCCCCTCCTGGAGCTGCAGCTCGTCCCCTTCTTCCTCTTGTTGCTCCTT 895

Qy        182 uSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAlaCy 202
              |||||||||:::                |||            :::   |||
Db        896 CTCCCCGGCCAGC--------------CCTGCCAGCTCCACCTCGGACAGTATTGAGTC 940
```

```
Qy        202 sLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                  |||      |||      :::   |||:::
Db        941 CTGGCAGCCCTCTGAGCTCCACCTGAACCCCTTCATG 977

RESULT 13
US-08-722-570-15
; Sequence 15, Application US/08722570
; Publication No. US20030044887A1
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  5365
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-15

Alignment Scores:
Pred. No.:              6.29e-19        Length:         1277
Score:                  304.00          Matches:        86
Percent Similarity:     57.29%          Conservative:   24
Best Local Similarity:  44.79%          Mismatches:     64
Query Match:            26.97%          Indels:         20
DB:                     8               Gaps:           8

US-09-595-947E-10 (1-214) x US-08-722-570-15 (1-1277)

Qy         24 AlaSerGluAspGluValThrCysProThrSerAlaProProSerProThrArgThrPro 43
              :::|||:::||||||                  :::|||   |||||||||   |||
Db        298 AGCTCGGACGATGAG---------CAGCTACACAGTCCGACAAGCCCGACGCTCACG--- 345

Qy         44 GlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLysLeuArgAla 63
                        :::::::|||     :::|||                    |||
Db        346 -----------CACCTGCAGCAGGGACGGGACCAGGGGGAGGAGAACAGCCCGCGATGC 393

Qy         64 ArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSer 83
              |||         |||||||||   :::   :::       |||   |||   :::::::::
Db        394 AGG---------AGGAGCCGAGCCCGCGGAGACACCGTGCTG---AAGATCAAGAAGACC 441

Qy         84 ArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeu 103
              ||||||   |||||||||:::|||||||||||||||||||||   ||||||  ||||||
Db        442 CGGCGCGTTAAAGCCAATAACCGCGAGAGGAATCGCATGCACCACCTGAACTATGCGCTC 501

Qy        104 AspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGlu 123
              |||:::||||||   |||||||||:::   |||:::||||||||||||||||||||||||
Db        502 GATTCTCTGAGGGAGGTTCTACCGTCATTACCCGAAGACGCCAAACTCACCAAGATAGAG 561

Qy        124 ThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAsp 143
              ||||||||||||||||||||||||||||||||||:::::|||||||||:::||||||
Db        562 ACCTTGCGCTTTGCCCACAACTACATCTGGGCTCTTAGCGAAACTTTGCGCCTGGCCGAC 621

Qy        144 ------HisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGlu------Leu 159
                  |||   :::|||        |||   :::   |||   |||          |||
Db        622 CAGCTGCACGG-ATCTAC---TTCCACCCCAGCAGCAGCCATATTGGTACAGGACTCCTA 677

Qy        160 GlySerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
                  |||      |||||||||                :::  |||
Db        678 TCCTTCCCTGA-GCCCCTCCTGGAGCTGCAGCTCGTCCCCATCCTCCAACTCTTGCGACT 736

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeu---LeuGlyAlaThr 198
                 |||      ||||||               |||      |||      ||:::
Db        737 CCTTCTCCCCGACCAGCCCTGCCAGCTCCACCTCGGACAGTATTGAGTACTGGCAGCCCT 796

Qy        199 SerSerAlaCysLeuSerProGlySerLeuAlaPhe 210
                 |||:::      :::|||      |||   |||
Db        797 CTGAGCTCCGCTTGAACCCCTTCATGTCTGCCCTTT 832

RESULT 14
US-10-004-717-43
; Sequence 43, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
;   LENGTH: 1550
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-43

Alignment Scores:
Pred. No.:              8.11e-13        Length:         1550
Score:                  238.50          Matches:        78
Percent Similarity:     39.29%          Conservative:   21
Best Local Similarity:  30.95%          Mismatches:     72
Query Match:            21.16%          Indels:         81
DB:                     14              Gaps:           8

US-09-595-947E-10 (1-214) x US-10-004-717-43 (1-1550)

Qy         19 ArgSerPheProArgAlaSerGluAsp------GluValThrCysProThrSerAlaPro 36
              |||  |||  |||  ||||||     ::::::      |||  |||  |||
Db        106 AGAAAGTTTGCTAGACAATGTGAGGACCAGAAACAAATTAAGAAACCAGAGAGCTTTCCA 165

Qy         37 --------------------ProSerProThrArgThrProGlyAsnCysAlaGluAla 49
                                   |||      |||  ||||||         |||
Db        166 AAACAAGTTGTCCTTCGAGGAAAGAGCATTAAAAGGGCCCCTGGAGAAGAAACCGAGAAA 225

Qy         50 Glu----------------------------GluGlyGlyCysArgGlyAlaProArg 59
              |||                          |||  |||
Db        226 GAAGAGGAGGAAGAAGACAGAGAGGAAGAAGATGAGAATGGC----------------- 267
```

```
Qy          60 LysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79
Db         268 ---TTGTCCAGAAGGAGGGGGCTCAGGAAAAAAAAGACCACCAAACTACGACTGGAAAGG 324

Qy          80 GlnArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99
Db         325 GTC---AAGTTCAGGAGACAGGAAGCTAATGCGCGCGAGAGGAACCGGATGCACGGCCTC 381

Qy         100 AsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeu 119
Db         382 AATGATGCTCTGGACAATTTGCGAAAAGTGGTCCCCTGTTACTCTAAAACCCAAAAACTG 441

Qy         120 ThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeu 139
Db         442 TCCAAAATAGAAACTTTACGACTGGCCAAAAATTACATCTGGGCACTTTCTGAAATTCTG 501

Qy         140 ArgIle---------------------------------------------------------- 141
Db         502 AGGATTGGCAAGAGACCGGATCTGCTCACGTTCGTCCAAAACTTATGCAAAGGTCTTTCC 561

Qy         141 ---------------------------------------------------------- 141
Db         562 CAGCCAACTACAAACTTGGTGGCAGGCTGCTTACAGCTCAACGCCAGAAGTTTCCTGATG 621

Qy         142 ------------------AlaAspHis------SerLeuTyrAlaLeuGluProProAla 153
Db         622 GGTCAGGGTGGGGAGGCTGCCCACCACACAAGGTCACCCTACTCCACATTCTACCCACCC 681

Qy         154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTrpGlySerLeuTyr 173
Db         682 TACCACAGCCCTGAGCTGGCCACTCCCCCAGGG----------------------CAT 717

Qy         174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgProGly 193
Db         718 GGGACTCTTGATAATTCCAAGTCCATGAAACCCTACAATTACTGCAGTGCATATGAATCC 777

Qy         194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerPro 205
Db         778 TTCTATGAAAGTACCTCCCCTGAGTGTGCCAGCCCT 813

RESULT 15
US-10-004-717-8
; Sequence 8, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 1957
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-8

Alignment Scores:
Pred. No.:              1.04e-12        Length:         1957
Score:                  238.50          Matches:        78
Percent Similarity:     39.29%          Conservative:   21
Best Local Similarity:  30.95%          Mismatches:     72
Query Match:            21.16%          Indels:         81

DB:                     14              Gaps:           8

US-09-595-947E-10 (1-214) x US-10-004-717-8 (1-1957)

Qy          19 ArgSerPheProArgAlaSerGluAsp------GluValThrCysProThrSerAlaPro 36
Db         182 AGAAAGTTTGCTAGACAATGTGAGGACCAGAAACAAATTAAGAAACCAGAGAGCTTTCCA 241

Qy          37 --------------------ProSerProThrArgThrProGlyAsnCysAlaGluAla 49
Db         242 AAACAAGTTGTCCTTCGAGGAAAGAGCATTAAAAGGGCCCCTGGAGAAGAAACCGAGAAA 301

Qy          50 Glu----------------------------GluGlyGlyCysArgGlyAlaProArg 59
Db         302 GAAGAGGAGGAAGAAGACAGAGAGGAAGAAGATGAGAATGGC---------------- 343

Qy          60 LysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79
Db         344 ---TTGTCCAGAAGGAGGGGGCTCAGGAAAAAAAAGACCACCAAACTACGACTGGAAAGG 400

Qy          80 GlnArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99
Db         401 GTC---AAGTTCAGGAGACAGGAAGCTAATGCGCGCGAGAGGAACCGGATGCACGGCCTC 457

Qy         100 AsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeu 119
Db         458 AATGATGCTCTGGACAATTTGCGAAAAGTGGTCCCCTGTTACTCTAAAACCCAAAAACTG 517

Qy         120 ThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeu 139
Db         518 TCCAAAATAGAAACTTTACGACTGGCCAAAAATTACATCTGGGCACTTTCTGAAATTCTG 577

Qy         140 ArgIle---------------------------------------------------------- 141
Db         578 AGGATTGGCAAGAGACCGGATCTGCTCACGTTCGTCCAAAACTTATGCAAAGGTCTTTCC 637

Qy         141 ---------------------------------------------------------- 141
Db         638 CAGCCAACTACAAACTTGGTGGCAGGCTGCTTACAGCTCAACGCCAGAAGTTTCCTGATG 697

Qy         142 ------------------AlaAspHis------SerLeuTyrAlaLeuGluProProAla 153
Db         698 GGTCAGGGTGGGGAGGCTGCCCACCACACAAGGTCACCCTACTCCACATTCTACCCACCC 757

Qy         154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTrpGlySerLeuTyr 173
Db         758 TACCACAGCCCTGAGCTGGCCACTCCCCCAGGG----------------------CAT 793

Qy         174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgProGly 193
Db         794 GGGACTCTTGATAATTCCAAGTCCATGAAACCCTACAATTACTGCAGTGCATATGAATCC 853

Qy         194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerPro 205
Db         854 TTCTATGAAAGTACCTCCCCTGAGTGTGCCAGCCCT 889

Search completed: February  2, 2004, 21:41:42
Job time : 361 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 2, 2004, 17:09:49 ; Search time 2044 Seconds
(without alignments)
2544.598 Million cell updates/sec

Title: US-09-595-947E-10
Perfect score: 1127
Sequence: 1 MTPQPSGAPTVQVTRETERS.....LGATSSACLSPGSLAFSDFL 214

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09595947/runat_02022004_154934_8059/app_query.fasta_1.391
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09595947_@CGN_1_1_2810_@runat_02022004_154934_8059 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	809	71.8	1540	11	AK008017	AK008017 Mus muscu
2	726	64.4	1025	14	BY708009	BY708009 BY708009
c 3	676.5	60.0	593	28	AZ296526	AZ296526 RPCI-23-1
4	340.5	30.2	932	14	CA979119	CA979119 AGENCOURT
5	327	29.0	865	14	CA496091	CA496091 AGENCOURT
6	327	29.0	1098	13	BQ169355	BQ169355 STR00303
7	318	28.2	603	14	CD282259	CD282259 G39271.16
8	317.5	28.2	632	12	BJ010277	BJ010277 BJ010277
9	314	27.9	687	14	CA945402	CA945402 UI-M-FD0-
10	311	27.6	604	13	BX308104	BX308104 BX308104
11	310.5	27.6	600	12	BG808248	BG808248 2083-52 M
12	299	26.5	600	13	BU924937	BU924937 7103-91 M
13	298.5	26.5	1001	9	AL540071	AL540071 AL540071
14	296.5	26.3	1039	10	BE780690	BE780690 601469349
15	295	26.2	730	13	BU612495	BU612495 UI-M-FR0-
c 16	285.5	25.3	1037	29	CNS03VJ9	AL262494 Tetraodon
17	281	24.9	823	13	BU054481	BU054481 UI-M-FD0-
18	280	24.8	629	12	BJ093114	BJ093114 BJ093114
19	278	24.7	595	12	BJ030202	BJ030202 BJ030202
20	267.5	23.7	401	9	AW147434	AW147434 da02h12.y
21	265	23.5	947	13	BX419330	BX419330 BX419330
22	251	22.3	588	9	AV673464	AV673464 AV673464
23	249	22.1	490	9	AV995230	AV995230 AV995230
24	249	22.1	711	13	BW275045	BW275045 BW275045
25	243	21.6	814	13	BQ178789	BQ178789 UI-M-EV0-
c 26	239	21.2	1003	29	CNS02IB1	AL198694 Tetraodon
27	234.5	20.8	1038	13	BX419494	BX419494 BX419494
28	231	20.5	1022	13	BU113216	BU113216 603129939
29	229.5	20.4	934	13	BU139911	BU139911 603132642
30	228	20.2	704	13	BU057851	BU057851 UI-M-FR0-
31	228	20.2	740	14	CA319439	CA319439 UI-M-FW0-
32	228	20.2	835	14	CA320553	CA320553 UI-M-FW0-
33	226.5	20.1	710	13	BU057238	BU057238 UI-M-FR0-
34	226.5	20.1	742	13	BQ572268	BQ572268 UI-M-FD0-
35	226.5	20.1	770	13	BQ572426	BQ572426 UI-M-FD0-
36	224.5	19.9	688	10	BG699059	BG699059 602678696
37	224.5	19.9	711	2	HSM067528	Bx478059 Homo sapi
38	224.5	19.9	901	13	BQ424098	BQ424098 AGENCOURT
39	224.5	19.9	1967	11	BC022560	BC022560 Homo sapi
40	222.5	19.7	953	9	AU067624	AU067624 AU067624
41	221.5	19.7	781	13	BU611678	BU611678 UI-M-FI0-
c 42	221.5	19.7	872	13	BX453565	BX453565 BX453565
43	219.5	19.5	724	10	BE783567	BE783567 601471617
44	219	19.4	722	2	HSM078917	Bx508253 Homo sapi
45	219	19.4	875	29	CNS02BKP	AL189970 Tetraodon

ALIGNMENTS

RESULT 1
AK008017
LOCUS AK008017 1540 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010001M19 product:atonal homolog 5 (Drosophila), full insert sequence.
ACCESSION AK008017
VERSION AK008017.1 GI:12841941
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1540)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.
FEATURES Location/Qualifiers
source 1. .1540
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:2010001M19"
/db_xref="MGI:1907403"
/db_xref="taxon:10090"
/clone="2010001M19"
/sex="male"
/tissue_type="small intestine"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 241. .885
/note="unnamed protein product; atonal homolog 5 (Drosophila) (MGD|MGI:893591, GB|NM_009719, evidence: BLASTN, 99%, match=648)
putative"
/codon_start=1
/protein_id="BAB25411.1"
/db_xref="GI:12841942"
/db_xref="MGI:893591"
/translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSHTLIPR DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKKRRSRRKKANDRERNRMHNLNSA LDALRGVLPTFPDDAKLTKVETLRFAHNYIWALTQTLRIADHSLYGPEPPVPCGELGS PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT 306 a 485 c 397 g 352 t
ORIGIN

Alignment Scores:

Pred. No.:	8.47e-46	Length:	1540
Score:	809.00	Matches:	161
Percent Similarity:	82.33%	Conservative:	16
Best Local Similarity:	74.88%	Mismatches:	36
Query Match:	71.78%	Indels:	2
DB:	11	Gaps:	2

US-09-595-947E-10 (1-214) x AK008017 (1-1540)

```
Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||       |||   |||:::||||||:::   ||||||::::::
Db        241 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 300

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||   ||||||:::   ||||||             |||   |||||||||   |||
Db        301 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCACACT 360

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                 |||   :::|||:::|||||||||   |||   |||||||||       ||||||
Db        361 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 420

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
```

```
                ||||||||||||||||||||||:::|||||||||||||||||||||||||||||:::  .
Db        421 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAAAAG 480

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
                |||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        481 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 540

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 600

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
                |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGGTCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 660

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
                ||||||||||||||||||||||      |||||||||   |||   ||||||||||||||
Db        661 ATAGCGGACCACAGCTTGTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 717

Qy        161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
                ||||||   ||||||      ||||||||||||||:::|||||||||||||||||||||
Db        718 AGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCG 777

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
                |||:::|||||||||   ||||||||||||||   |||||||||          :::|||
Db        778 GGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCC 837

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                :::   |||   ||||||:::|||   ||||||||||||||
Db        838 CCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 882
```

RESULT 2
BY708009
LOCUS BY708009 1025 bp mRNA linear EST 16-DEC-2002
DEFINITION BY708009 RIKEN full-length enriched, adult male small intestine Mus musculus cDNA clone 2010001M19 5', mRNA sequence.
ACCESSION BY708009
VERSION BY708009.1 GI:27119192
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1025)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .1025
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="2010001M19"
/sex="male"
/tissue_type="small intestine"
/dev_stage="adult"
/lab_host="SOLR"
/clone_lib="RIKEN full-length enriched, adult male small intestine"
/note="Site_1: XhoI; Site_2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI."
BASE COUNT 215 a 316 c 286 g 207 t 1 others
ORIGIN

Alignment Scores:

Pred. No.:	2.75e-40	Length:	1025
Score:	726.00	Matches:	160
Percent Similarity:	82.33%	Conservative:	17
Best Local Similarity:	74.42%	Mismatches:	36
Query Match:	64.42%	Indels:	7

```
DB:                 14          Gaps:          2

US-09-595-947E-10 (1-214) x BY708009 (1-1025)

Qy          1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
              |||   |||   |||     |||   |||:::||||||:::   ||||||::::::
Db        241 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 300

Qy         21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
              ||||||  ||||||:::  ||||||          |||  |||||||||   |||
Db        301 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCACACT 360

Qy         41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                 |||   :::|||:::||||||||||  |||  |||||||||     ||||||
Db        361 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 420

Qy         61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||:::
Db        421 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAAAAG 480

Qy         81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
              ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        481 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 540

Qy        101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 600

Qy        121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGGTCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 660

Qy        141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
              ||||||||||||||||||      |||||||||   |||   ||||||||||||||||
Db        661 ATAGCGGACCACAGCTTGTATGGCCCGGAGCCCCCTGTGCCC---TGTGGAGAGCTGGGG 717

Qy        161 SerProGly---GlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
              ||||||   |||      |||||||||||||:::||||||   ||||||||||||
Db        718 AC-CCCGGACGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCTCAGTCTCCCAAGCG 776

Qy        180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
              |||:::||||||||||   ||||||||||||:::   ||||||||   ||||||:::|||
Db        777 GGTAACCTGAGCCCACG-GCCTCATTGGAGGATTA-CCTGGCCTGCA-GGTGCCAGCTCC 833

Qy        200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
                 :::  |||   ||||||:::|||   ||||||||||||||
Db        834 CCATCTAT-CTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTG 877

RESULT 3
AZ296526/c
LOCUS       AZ296526                593 bp    DNA     linear   GSS 27-JUL-2000
DEFINITION  RPCI-23-160G18.TV RPCI-23 Mus musculus genomic clone RPCI-23-160G18
            , genomic survey sequence.
ACCESSION   AZ296526
VERSION     AZ296526.1  GI:9538311
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI-23-160G18.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 160  row: G  column: 18
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .593
                      /organism="Mus musculus"
                      /mol_type="genomic DNA"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="RPCI-23-160G18"
                      /sex="Female"
                      /lab_host="DH10B"
                      /clone_lib="RPCI-23"
                      /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                      EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                      brain genomic DNA was isolated and  partially digested
                      with a combination of EcoRI and EcoRI Methylase.   Size
                      selected DNA was cloned into the pBACe3.6 vector at the
                      EcoRI  sites.  The ligation products were transformed into
                      DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT       88 a    159 c    213 g    133 t
ORIGIN

Alignment Scores:
Pred. No.:              3.96e-37        Length:         593
Score:                  676.50          Matches:        136
Percent Similarity:     82.68%          Conservative:   12
Best Local Similarity:  75.98%          Mismatches:     30
Query Match:            60.03%          Indels:         2
DB:                     28              Gaps:           1

US-09-595-947E-10 (1-214) x AZ296526 (1-593)

Qy          3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
              |||   |||       |||   |||:::||||||:::   ||||||::::::   ||||||
Db        547 CCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCC 488

Qy         23 ArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThrArgThr 42
                ||||||:::  ||||||          |||  ||||||||||||||||
Db        487 GGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCATA 428

Qy         43 ProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLysLeuArg 62
              |||   :::|||:::||||||||||  |||  |||||||||     ||||||||||||
Db        427 CCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGC 368

Qy         63 AlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArg 82
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        367 GCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAGCGAAGA 308

Qy         83 SerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAla 102
              ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        307 AGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCG 248

Qy        103 LeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIle 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 188

Qy        123 GluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAla 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCG 128

Qy        143 AspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySerPro 162
              |||||||||   |||      |||||||||   |||      ||||||||||||||||
Db        127 GACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCCTG----GGAGAGCTGGGGAGCCCC 72

Qy        163 GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGlySer 181
```

```
              ||||||   |||         |||:::   :::|||||||||||||||||||||:::
Db         71 GGAGGGGCTCCAACGGGGACTGGGGCTCATCACTCCCCAGTCTCCCAAGCGGGAACC 15

RESULT 4
CA979119
LOCUS       CA979119                932 bp    mRNA    linear   EST 06-JAN-2003
DEFINITION  AGENCOURT_11295215 NIH_MGC_164 Mus musculus cDNA clone
            IMAGE:30146192 5', mRNA sequence.
ACCESSION   CA979119
VERSION     CA979119.1  GI:27511773
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe and Dr. Mina
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM0061  row: h  column: 09
            High quality sequence start: 16
            High quality sequence stop: 640.
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30146192"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_164"
                     /note="Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2:
                     NotI; Non-normalized full-length enriched library from
                     pooled mouse embryonic limb, maxilla and mandible, day
                     10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
                     Cloned directionally, priming method: Oligo-dT.  cDNA
                     enrichment: >1k bp, Average insert size 1.8k bp. Priming
                     sequence: 5'GACTAGTTCTAGATCGCGAGCGGCCGCCC(T) 3'. Tissue
                     contributed by, David Rowe. Library constructed by ResGen,
                     Invitrogen Corp."
BASE COUNT      153 a    349 c    276 g    153 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              6.1e-14         Length:         932
Score:                  340.50          Matches:        103
Percent Similarity:     50.00%          Conservative:   21
Best Local Similarity:  41.53%          Mismatches:     64
Query Match:            30.21%          Indels:         61
DB:                     14              Gaps:           11

US-09-595-947E-10 (1-214) x CA979119 (1-932)

Qy          4 GlnProSerGlyAlaProThrValGlnValThr---ArgGluThrGluArgSerPhePro 22
              :::|||:::     |||   |||::::::::::       |||   |||     |||
Db        208 AGGCCGACCTCCAAACCTCCTGTCCGTCTGTCGGTCCTGCACACTGCAAGATGCCTGCCC 267

Qy         23 --------------------------------ArgAlaSerGluAspGluValThrCys 31
                                              |||:::            ||||||
Db        268 CTTTGGAGACCTGCATCTCTGATCTCGACTGCTCCAGCAGCAACAGCAGCAGCGACCTGT 327

Qy         32 ProThrSerAlaProProSerProThrArgThrProGlyAsnCysAlaGluAlaGluGlu 51
              |||   |||:::|||         |||      ||||||           |||
Db        328 CCAGCTTCCTCACCGACGAGGAGGACT---GTGCCAGGCTACAGCCCCTA-GCCTCCACC 383

Qy         52 GlyGlyCys--------------ArgGlyAlaPro----------------------- 58
                 |||                 |||   ||||||
Db        384 TCGGGGCTGTCCGTGCCAGCCCGGAGGAGCGCTCCCGCCCTCTCCGGGGCATCGAATGTT 443

Qy         59 ------------------ArgLysLeuArgAlaArgArgGlyGlyArgSerArgProLys 72
                                  :::   |||   |||||   ||||||:::|||   :::
Db        444 CCCGGTGCCCAGGACGAAGAGCAGGAACGGCGGAGGCGGCGAGGTCGCGCTCGGGTGCGG 503

Qy         73 SerGluLeuAlaLeuSerLysGlnArgArgSerArgArgLysLysAlaAsnAspArgGlu 92
              ||||||      |||         ||||||||||||||   ||||||||||||||||||
Db        504 TCCGAGGCTCTGCTGCACTCCCTGCGGAGGAGTCGTCGCGTCAAAGCCAACGATCGCGAG 563

Qy         93 ArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuProThr 112
              ||||||||||||||:::||||||:::||||||||||||||||||    |||||||||:::
Db        564 CGCAACCGCATGCACAACCTCAACGCTGCGCTGGACGCCTTGCGCAGCGTGCTGCCCTCG 623

Qy        113 PheProAspAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHis-AsnTyrIl 132
              ||||||||||   |||||||||||||||||||||||||||||:::|||||||
Db        624 TTCCCCGACGACACCAAGCTCACCAAGATTGAGACGCTGCGCTTCGCCTACAAACTACAT 683

Qy        132 eTrp-AlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeuTyrAlaLeuGluProP 152
              |||| ||||||   :::|||||||||:::||||||      |||                |
Db        684 CTGGGGCCCTGGCTGAGACACTGCGCCTGGCAGATCAAGGGCTC--------------C 728

Qy        152 roAlaProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTrpGlySerL 172
              ||                        :::||||||   ||||||           |||
Db        729 CCGGGGGC--------------AGTGCCCCGGGAGCGCCTCCT---------GCCTCCG 764

Qy        172 euTyrSerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgP 192
                    |||:::|||                        |||   :::   |||   ||||
Db        765 CAGTGTGTCCCCTGTCT--------------------GCCCGGGCCCCCGAGCCCCGGC 803

Qy        192 roGlyLeuLeu------GlyAlaThrSerSerAlaCysLeuSerPro------------G 206
              ||      |||      |||   :::   :::   |||      |||            |
Db        804 CAGCGACACTGAGTCCTGGGGGTTCCGGGGCNCGCTGCCCTCCCCCTTGCGGCCACTGGG 863

Qy        206 lySerLeuAlaPheSerAsp 212
              ||          |||||||||
Db        864 GGCATCACCACTTCTCTGAC 883

RESULT 5
CA496091
LOCUS       CA496091                865 bp    mRNA    linear   EST 14-NOV-2002
DEFINITION  AGENCOURT_10812072 NCI_CGAP_ZEmb2 Danio rerio cDNA clone
            IMAGE:6789795 5', mRNA sequence.
ACCESSION   CA496091
VERSION     CA496091.1  GI:24959169
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 865)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Leonard I. Zon, M.D.
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14290  row: n  column: 02
            High quality sequence stop: 658.
FEATURES             Location/Qualifiers
     source          1. .865
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
```

```
                     /clone="IMAGE:6789795"
                     /tissue_type="embryo"
                     /lab_host="DH10B (T1-resistant)"
                     /clone_lib="NCI_CGAP_ZEmb2"
                     /note="Vector: pCMV-SPORT6.ccdb; Site_1: EcoRV; Site_2:
                     NotI; Cloned unidirectionally. Primer: Oligo dT. Average
                     insert size 2 kb. Constructed by J. Wang (Research
                     Genetics, Invitrogen Corp) from tissue donated by L. Zon
                     (Harvard University). Note: this is a NCI_CGAP Library."
BASE COUNT      226 a    256 c    214 g    169 t
ORIGIN

Alignment Scores:
Pred. No.:              4.77e-13        Length:         865
Score:                  327.00          Matches:        84
Percent Similarity:     53.14%          Conservative:   26
Best Local Similarity:  40.58%          Mismatches:     65
Query Match:            29.02%          Indels:         32
DB:                     14              Gaps:           6

US-09-595-947E-10 (1-214) x CA496091 (1-865)

Qy          3 ProGlnProSerGlyAlaProThrValGlnVal---------------------------- 13
              ||||||   |||   :::|||   ::::::::
Db         85 CCTCAAGCATCTCCCAGCCCACCAATAAGGTTATCAACAATGGAGATCGTATACTCCGAT 144

Qy         14 ------ThrArgGluThrGluArgSerPheProArgAlaSerGluAspGluValThrCys 31
                  |||         :::   ||||||         ::::::||||||   :::
Db        145 ATGGAAACCTCAAGCTGTGACTACTCCTTTTCGCAC---ACGGATGATGAAGACTCGCGC 201

Qy         32 ProThrSerAlaProProSerProThrArgThrProGlyAsnCysAlaGluAlaGluGlu 51
                :::     |||   ||||||               :::|||
Db        202 AGCAGCCTCCACCCCGCGTCCCCCGCG---------TCCTCCTGC--------------- 237

Qy         52 GlyGlyCysArgGlyAlaProArgLysLeuArgAlaArgArgGlyGlyArgSerArgPro 71
              |||           :::|||      |||:::   ::::::      |||   |||
Db        238 GGAAAACCACCTGCGTCTCCAGCCGGGCTCCAGCAGAAGAAAAGGCGCAGGGGGCGCGCG 297

Qy         72 LysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArgLysLysAlaAsnAspArg 91
              ::::::|||      :::         :::::::::||||||  ||||||||||||||
Db        298 AGGAACGAAACCACTGTGCACGTCGTGAAGAAGAACCGCAGGCTGAAGGCCAACGACCGC 357

Qy         92 GluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuPro 111
              ||||||||||||||||||:::||||||  ||||||||||||||||||   |||||||||
Db        358 GAGAGGAACAGGATGCACAACCTTAACGACGCATTGGATGCTTTGAGAAGCGTCCTGCCT 417

Qy        112 ThrPheProAspAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHisAsnTyr 131
              ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db        418 GCGTTTCCTGACGACACAAAGCTGACCAAAATTGAGACTCTGCGCTTCGCTCACAACTAC 477

Qy        132 IleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu--------------- 146
              ||||||||||||:::::|||:::||||||||||||
Db        478 ATCTGGGCACTTTCGGAGACCATCCGGATCGCAGACCAGAAGCAGGGCAAGTCAAGAGAC 537

Qy        147 ---TyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyPro 165
                 |||  |||                  :::   |||||||||
Db        538 GGTCCGCTGCTGCTCCCCGGACTAAGCTGCATGGCAGATGCACCCAGCCCCGGCAGTGAC 597

Qy        166 ProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGlySerLeu--------- 182
                 |||  |||      |||      |||  :::  |||
Db        598 TCTTGCTCCTGGCCGTCGGGGGCATCCTCGTCGTCTTCATCACCGTCTTACTGCAACTCA 657

Qy        183 ---------SerProAlaAla 186
                      ||||||||||||
Db        658 GACCCGGGCAGCCCCGCAGCC 678

RESULT 6
BQ169355
LOCUS       BQ169355                1098 bp    mRNA    linear   EST 01-MAY-2002
DEFINITION  STR00303 segmentation stage cDNA library Danio rerio cDNA clone
            CB260 5' similar to NEUROGENIN 1, mRNA sequence.
ACCESSION   BQ169355
VERSION     BQ169355.1  GI:20376783
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 1098)
  AUTHORS   Loppin,B., Pflumio,S., Steffan,T., Heyer,V., Furthauer,M., Thisse
            ,C. and Thisse,B.
  TITLE     Expression of the zebrafish genome during embryogenesis (2002)
  JOURNAL   Unpublished
COMMENT     Contact: Thisse B
            Institut de Genetique et de Biologie Moleculaire et Cellulaire
            CNRS, INSERM, ULP
            1, rue Laurent Fries, BP163, CU de Strasbourg, 67404 Illkirch Cedex
            , France
            Tel: 33 3 88 65 33 60
            Fax: 33 3 88 65 32 01
            Email: thisse@igbmc.u-strasbg.fr
            EST from a cDNA of a gene whose expression is spatially restricted
            during embryogenesis. We have established its expression pattern
            during embryonic development by whole mount in situ hybridization
            on zebrafish embryos from the gastrula stage to 2 days of
            development. The corresponding data are available on the zebrafish
            community database at http://zfin.org/ cDNA library preparation: B.
            Riggleman. DNA Sequencing by:IGBMC sequencing facility. Clone
            distribution: zebrafish international resource center at the
            University of Oregon (Institute of Neuroscience, 1254 University of
            Oregon, Eugene, OR 97403-1254)
            Seq primer: T3 ATTAACCCTCACTAAAGGGA.
FEATURES             Location/Qualifiers
     source          1. .1098
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="CB260"
                     /dev_stage="segmentation stage embryos"
                     /clone_lib="segmentation stage cDNA library"
                     /note="Vector: Lambda Zap; Site_1: EcoRI; Site_2: XhoI;
                     Oligo dT cDNA library constructed from RNA pooled from
                     segmentation stage zebrafish embryos"
BASE COUNT      284 a    296 c    264 g    254 t
ORIGIN

Alignment Scores:
Pred. No.:              6e-13           Length:         1098
Score:                  327.00          Matches:        84
Percent Similarity:     53.14%          Conservative:   26
Best Local Similarity:  40.58%          Mismatches:     65
Query Match:            29.02%          Indels:         32
DB:                     13              Gaps:           6

US-09-595-947E-10 (1-214) x BQ169355 (1-1098)

Qy          3 ProGlnProSerGlyAlaProThrValGlnVal---------------------------- 13
              ||||||   |||   :::|||   ::::::::
Db        115 CCTCAAGCATCTCCCAGCCCACCAATAAGGTTATCAACAATGGAGATCGTATACTCCGAT 174

Qy         14 ------ThrArgGluThrGluArgSerPheProArgAlaSerGluAspGluValThrCys 31
                  |||         :::   ||||||         ::::::||||||   :::
Db        175 ATGGAAACCTCAAGCTGTGACTACTCCTTTTCGCAC---ACGGATGATGAGGACTCGCGC 231

Qy         32 ProThrSerAlaProProSerProThrArgThrProGlyAsnCysAlaGluAlaGluGlu 51
                :::     |||   ||||||               :::|||
Db        232 AGCAGCCTCCACCCCGCGTCCCCCGCG---------TCCTCCTGC--------------- 267

Qy         52 GlyGlyCysArgGlyAlaProArgLysLeuArgAlaArgArgGlyGlyArgSerArgPro 71
              |||           :::|||      |||:::   ::::::      |||   |||
Db        268 GGAAAACCACCTGCGTCTCCAGCCGGGCTCCAGCAGAAGAAAAGGCGCAGGGGGCGCGCG 327

Qy         72 LysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArgLysLysAlaAsnAspArg 91
              ::::::|||      :::         :::::::::||||||  ||||||||||||||
Db        328 AGGAACGAAACCACTGTGCACGTCGTGAAGAAGAACCGCAGGCTGAAGGCCAACGACCGC 387
```

```
Qy         92 GluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuPro 111
              ||||||||||||||||||:::||||||   ||||||||||||||||||   |||||||||
Db        388 GAGAGGAACAGGATGCACAACCTTAACGACGCATTGGATGCTTTGAGAAGCGTCCTGCCT 447

Qy        112 ThrPheProAspAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHisAsnTyr 131
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        448 GCGTTTCCTGACGACACAAAGCTGACCAAAATTGAGACTCTGCGCTTCGCTCACAACTAC 507

Qy        132 IleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu---------------- 146
              ||||||||||||:::::::|||:::||||||||||||
Db        508 ATCTGGGCACTTTCGGAGACCATCCGGATCGCAGACCAGAAGCAGGGCAAGTCAAGAGAC 567

Qy        147 ---TyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyPro 165
                 |||   |||                    :::      |||||||||
Db        568 GGTCCGCTGCTGCTCCCCGGACTAAGCTGCATGGCAGATGCACCCAGCCCCGGCAGTGAC 627

Qy        166 ProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGlySerLeu--------- 182
                 |||   |||      |||      |||   :::   |||
Db        628 TCTTGCTCCTGGCCGTCGGGGGCATCCTCGTCGTCTTCATCACCGTCTTACTGCAACTCA 687

Qy        183 ---------SerProAlaAla 186
                      ||||||||||
Db        688 GACCCGGGCAGCCCCGCAGCC 708
```

RESULT 7
CD282259
LOCUS CD282259 603 bp mRNA linear EST 23-MAY-2003
DEFINITION G39271.16 NCI_CGAP_ZEmb2 Danio rerio cDNA clone IMAGE:6521499 5', mRNA sequence.
ACCESSION CD282259
VERSION CD282259.1 GI:31060035
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 603)
AUTHORS Amundsen,C., Cachuela,N., Chen,F., Cheung,L.M., Chong,A., Murray,L., Oliva,J., Park,C., Reyes,J., Yungen,J. and Swimmer,C.
TITLE Expressed sequence tags from NCI_CGAP_ZEmb2, a Danio rerio embryonic library
JOURNAL Unpublished
COMMENT Contact: Chen F.
Exelixis, Inc.
170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA
Tel: 650 837 7000
Fax: 650 837 8300
Email: fchen@exelixis.com
DNA Sequencing by: Exelixis, Inc. Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: 14108 row: O column: 3
High quality sequence stop: 603.
FEATURES Location/Qualifiers
source 1. .603
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:6521499"
/tissue_type="embryo"
/lab_host="DH10B (T1-resistant)"
/clone_lib="NCI_CGAP_ZEmb2"
/note="Vector: pCMV-SPORT6.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Constructed by J. Wang (Research Genetics, Invitrogen Corp) from tissue donated by L. Zon (Harvard University). Note: this is a NCI_CGAP Library."
BASE COUNT 159 a 185 c 155 g 104 t
ORIGIN

Alignment Scores:

Pred. No.: 1.39e-12 Length: 603
Score: 318.00 Matches: 76
Percent Similarity: 54.89% Conservative: 25
Best Local Similarity: 41.30% Mismatches: 57
Query Match: 28.22% Indels: 26
DB: 14 Gaps: 5

US-09-595-947E-10 (1-214) x CD282259 (1-603)

```
Qy          3 ProGlnProSerGlyAlaProThrValGlnVal--------------------------- 13
              ||||||   |||   :::|||   :::::::::
Db         79 CCTCAAGCATCTCCCAGCCCACCAATAAGGTTATCAACAATGGAGATCGTATACTCCGAT 138

Qy         14 ------ThrArgGluThrGluArgSerPheProArgAlaSerGluAspGluValThrCys 31
                    |||          :::   ||||||         ::::::|||||   :::
Db        139 ATGGAAACCTCAAGCTGTGACTACTCCTTTTCGCAC---ACGGATGATGAGGACTCGCGC 195

Qy         32 ProThrSerAlaProProSerProThrArgThrProGlyAsnCysAlaGluAlaGluGlu 51
              :::      |||   ||||||                :::|||
Db        196 AGCAGCCTCCACCCCGCGTCCCCCGCG---------TCCTCCTGC--------------- 231

Qy         52 GlyGlyCysArgGlyAlaProArgLysLeuArgAlaArgArgGlyGlyArgSerArgPro 71
              |||            :::|||      |||:::   ::::::      |||   |||
Db        232 GGAAAACCACCTGCGTCTCCAGCCGGGCTCCAGCAGAAGAAAAGGCGCAGGGGGCGCGCG 291

Qy         72 LysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArgLysLysAlaAsnAspArg 91
              ::::::|||      :::        :::::::::||||||  ||||||||||||||
Db        292 AGGAACGAAACCACTGTGCACGTCGTGAAGAAGAACCGCAGGCTGAAGGCCAACGACCGC 351

Qy         92 GluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuPro 111
              ||||||||||||||||||:::||||||   ||||||||||||||||||   |||||||||
Db        352 GAGAGGAACAGGATGCACAACCTTAACGACGCATTGGATGCTTTGAGAAGCGTCCTGCCT 411

Qy        112 ThrPheProAspAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHisAsnTyr 131
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        412 GCGTTTCCTGACGACACAAAGCTGACCAAAATTGAGACTCTGCGCTTCGCTCACAACTAC 471

Qy        132 IleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu---------------- 146
              ||||||||||||:::::::|||:::||||||||||||
Db        472 ATCTGGGCACTTTCGGAGACCATCCGGATCGCAGACCAGAAGCAGGGCAAGTCAAGAGAC 531

Qy        147 ---TyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyPro 165
                 |||   |||                    :::      |||||||||
Db        532 GGTCCGCTGCTGCTCCCCGGACTAAGCTGCATGGCAGATGCACCCAGCCCCGGCAGTGAC 591

Qy        166 ProGlyAspTrp 169
                 |||
Db        592 TCTTGCTCCTGG 603
```

RESULT 8
BJ010277
LOCUS BJ010277 632 bp mRNA linear EST 05-DEC-2001
DEFINITION BJ010277 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA142H01 5', mRNA sequence.
ACCESSION BJ010277
VERSION BJ010277.1 GI:17364159
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha; Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 632)
AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855

```
          Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA142H01"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
BASE COUNT      120 a    231 c    190 g     90 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              1.58e-12        Length:         632
Score:                  317.50          Matches:        91
Percent Similarity:     50.45%          Conservative:   20
Best Local Similarity:  41.36%          Mismatches:     55
Query Match:            28.17%          Indels:         55
DB:                     12              Gaps:           11

US-09-595-947E-10 (1-214) x BJ010277 (1-632)

Qy          2 ThrProGln-----------ProSerGlyAlaProThrValGlnValThrArgGluThr 17
              |||||||||           |||:::  |||  |||   ::::::          |||
Db         92 ACCCCCCAGATTGGAACCCGTCCCGCTGCGGCATGGACTCCGCGTATTCGGACATCGACA 151

Qy         18 GluArgSer---------ProProArgAlaSerGluAspGluValThrCysProThrSer 34
                  |||          |||           |||::::::        |||
Db        152 GC-AACAGCTGCGACTTTTTCTCACACACTGACGAGGAGGACGCCCGAGTGCCG------ 204

Qy         35 AlaProProSerProThrArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCys 54
                 |||   ||||||  |||
Db        205 CGCCCCACCTCTCCTGAGCGC---------------------------------------- 225

Qy         55 ArgGlyAlaProArgLysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGlu 74
                       ::::::   |||   ||||||   ||||||:::|||          |||
Db        226 ---------GGGCAACAGCAAAGGAAGCGGCGCCGAGGCCGGGCGCGC------TGCGAG 270

Qy         75 LeuAlaLeuSerLysGlnArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsn 94
                 :::         :::::::::||||||  ||||||||||||||||||||||||
Db        271 GCGGTCGTGCAGGTGGTGAAAAAGAACCGGCGGCTGAAGGCTAACGACCGGGAACGGAAC 330

Qy         95 ArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPhePro 114
              |||||||||:::||||||   |||||||||   ||||||||||||||||||   ||||||
Db        331 CGCATGCATAACCTGAACGACGCGCTGGATGAGCTGCGCGGCGTCCTGCCGGCCTTTCCG 390

Qy        115 AspAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAla 134
              |||:::  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        391 GACGAAACCAAGCTGACCAAGATCGAGACTCTGCGCTTTGCGCACAACTACATCTGGGCT 450

Qy        135 LeuThrGlnThrLeuArgIleAlaAspHisSerLeuTyrAla---------LeuGluPro 151
              |||:::::::|||:::||||||||||||      |||   |||              :::
Db        451 CTGTCCGAGACCATCCGCATCGCAGAC------CTGCAGGCGGGGAGGCGCAACGACACC 504

Qy        152 ProAla---------ProHisCys---GlyGluLeuGlySerProGlyGlyProProGly 167
              |||             |||  |||     |||      ||||||        |||
Db        505 CCTCTGCTGCTTAACCCCTCGTGCCTCCGCGAGACCCCGAGTCCAGCCAGTGACCCCTGC 564

Qy        168 AspTrpGlySerLeuTyrSerProValSerGlnAlaGlySerLeuSerProAlaAlaSer 187
              |||             ||||||          :::   :::   |||||||||
Db        565 ACGTGG------------AGCCCCAGCGGATCCTCGTCCGCCTCCTCCCCGGCG------ 606

Qy        188 LeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAlaCysLeuSerProGlySer 207
                                                                |||||||||||
Db        607 ----------------------------------TACTGCGCGTNCAGCCCGGGCAGC 630

RESULT 9
CA945402
LOCUS       CA945402                687 bp    mRNA    linear   EST 30-DEC-2002
DEFINITION  UI-M-FD0-cdh-1-12-0-UI.r1 NIH_BMAP_FD0 Mus musculus cDNA clone
            IMAGE: 6828925 5', mRNA sequence.
ACCESSION   CA945402
VERSION     CA945402.1  GI:27433882
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6828925"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FD0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     , is TGAGAGAGCC. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instututes of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      117 a    228 c    250 g     91 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              2.96e-12        Length:         687
Score:                  314.00          Matches:        83
Percent Similarity:     54.50%          Conservative:   20
Best Local Similarity:  43.92%          Mismatches:     28
Query Match:            27.86%          Indels:         58
DB:                     14              Gaps:           9

US-09-595-947E-10 (1-214) x CA945402 (1-687)

Qy         16 GluThrGluArgSerPheProArgAla-----------------SerGluAspGluVal 29
              |||      |||   |||   ::::::                      |||:::||||||
Db         77 GAGCCGCGTAGGATGTTCGTCAAATCTGAGACTCTGGAGTTGAAGGAGGAAGAGGAGGTA 136

Qy         30 -----------------ThrCysProThrSerAlaProProSerPro------------ 39
                                     |||   ||||||      :::|||
Db        137 CTGATGCTGCTGGGCTCGGCTTCCCCGGCCTCGGCGACCCTGACCCCGATGTCCTCCAGC 196

Qy         40 ------------------------ThrArgThrProGlyAsnCys-------------- 46
                                        |||   ||||||:::
```

```
Db        197 GCGGACGAGGAGGAGGACGAGGAGCTGCGCCGGCCGGGCTCCGCGCGTGGGCAGCGTGGA 256

Qy         47 AlaGluAlaGluGlu----------------------------GlyGlyCysArgGly 56
              |||||||||   :::                            ||||||||||||
Db        257 GCGGAAGCCGGGCAGGGGGTGCAGGGCAGTCCGGCGTCGGGTGCCGGGGGTTGCCGG--- 313

Qy         57 AlaProArgLysLeuArgAla------------------ArgArgGlyGlyArgSerArg 70
                 |||   :::|||                       ||||||      |||||||||
Db        314 ---CCAGGGCGGCTGCTGGGCCTGATGCACGAGTGCAAGCGTCGCCCGTCGCGCTCACGG 370

Qy         71 ---------------ProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArg 85
                           |||:::         :::   :::   :::::::::||||||
Db        371 GCCGTCTCCCGAGGTGCCAAGACGGCGGAGACGGTGCAGCGCATCAAGAAGACCCGCAGG 430

Qy         86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
              |||||||||:::||||||||||||||||||||:::||||||:::||||||||||||
Db        431 CTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCGCTGGACGCG 490

Qy        106 LeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeu 125
              ||||||   ||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        491 CTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATCGAGACGCTG 550

Qy        126 ArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSer 145
              ||||||||||||||||||||||||||||||:::||||||||:::||||||   :::
Db        551 CGCTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCGGACACTGCG 610

Qy        146 LeuTyrAlaLeuGluProProAlaPro 154
                           |||||||||
Db        611 ------------------CCGGCGCCG 619
```

RESULT 10
BX308104
LOCUS BX308104 604 bp mRNA linear EST 07-APR-2003
DEFINITION BX308104 AGENAE Rainbow trout multi-tissues subtracted library (tcay) Oncorhynchus mykiss cDNA clone tcay0016b.k.15 5prim, mRNA sequence.
ACCESSION BX308104
VERSION BX308104.1 GI:29588749
KEYWORDS EST.
SOURCE Oncorhynchus mykiss (rainbow trout)
ORGANISM Oncorhynchus mykiss
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE 1 (bases 1 to 604)
AUTHORS Govoroun,M., Guiguen,Y. and Le Gac,F.
TITLE Construction and primary characterization of normalized cDNA libraries in rainbow trout, Oncorhynchus mykiss
JOURNAL Unpublished
COMMENT Contact: Guiguen Y
INRA - SCRIBE
Campus de beaulieu, RENNES cedex, 35042, France
Tel: 02.23.48.50.09
Fax: 02.23.48.50.20
Email: Yann.Guiguen@beaulieu.rennes.inra.fr
Sequence cleaned of vector, adaptator and repetitions. Contact us at sigenasupport@jouy.inra.fr to obtain the chromatogram of this sequence.
Plate: 0016 row: k column: 15
Seq primer: M13F.
FEATURES Location/Qualifiers
source 1. .604
/organism="Oncorhynchus mykiss"
/mol_type="mRNA"
/db_xref="taxon:8022"
/clone="tcay0016b.k.15"
/tissue_type="adipose tissue, blood, brain, differentiating gonads, gills, interrenal, intestine, kidney, liver, muscle, ovary, pituitary, testis"
/dev_stage="from embryos to adults"
/lab_host="DH10B"
/clone_lib="AGENAE Rainbow trout multi-tissues subtracted library (tcay)"
/note="Vector: pT7T3D-pac; Clone distribution : AGENAE Resource centre. Francois PIUMI, Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex, FRANCE"
BASE COUNT 162 a 178 c 148 g 116 t
ORIGIN

Alignment Scores:

Pred. No.:	4.2e-12	Length:	604
Score:	311.00	Matches:	64
Percent Similarity:	68.60%	Conservative:	19
Best Local Similarity:	52.89%	Mismatches:	32
Query Match:	27.60%	Indels:	6
DB:	13	Gaps:	2

US-09-595-947E-10 (1-214) x BX308104 (1-604)

```
Qy         26 GluAspGluValThrCysProThr---------SerAlaProProSerProThrArgThr 42
              :::||||||   :::|||   :::          |||:::|||   |||
Db        203 GATGACGAAGATTCGTGCAGCAGCATGCACGCTTCCTCCCCTTCATCCTCCATCGGCAAG 262

Qy         43 ProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLysLeuArg 62
              |||   :::   |||          :::|||   |||                :::   :::
Db        263 CCCGCCTCTCCGGCCAGCGACAGCCAGGGCTCGTGCTCTACGGATCTGGCACAGAAGAAG 322

Qy         63 AlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArg 82
                 |||||||||          |||   ::::::|||      :::         ::::::
Db        323 AGACGCAGAGGC---------AGAGCGAGGAACGAAGCCACCGTTCACGTGGTAAAGAAG 373

Qy         83 SerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAla 102
              :::||||||   ||||||||||||||||||||||||||||||||||   ||||||   |||
Db        374 AACCGGCGCGTAAAGGCAAATGACCGCGAGAGGAACAGAATGCACAGCCTGAATGACGCG 433

Qy        103 LeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIle 122
              |||:::   ||||||   |||||||||   ||||||||||||:::|||||||||||||||
Db        434 TTGGAGACCCTCCGGACCGTTCTACCGGCGTTCCCCGATGACAGCAAACTCACCAAGATC 493

Qy        123 GluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAla 142
              ||||||||||||||||||||||||||||||||||||||:::::::|||:::|||||||||
Db        494 GAAACTCTGCGCTTCGCTCACAATTACATCTGGGCACTCTCCGAGACCATACGCATCGCG 553

Qy        143 Asp 143
              |||
Db        554 GAT 556
```

RESULT 11
BG808248
LOCUS BG808248 600 bp mRNA linear EST 20-DEC-2001
DEFINITION 2083-52 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA, mRNA sequence.
ACCESSION BG808248
VERSION BG808248.1 GI:17955225
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 600)
AUTHORS Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W., White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
TITLE Gene expression in the developing mouse retina by EST sequencing and microarray analysis
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)
MEDLINE 21671825
PUBMED 11812828
COMMENT Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.
FEATURES Location/Qualifiers

```
     source          1. .600
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="neural retina"
                     /dev_stage="embryonic day 14.5 post-fertilization"
                     /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
BASE COUNT       94 a    238 c    161 g    107 t
ORIGIN

Alignment Scores:
Pred. No.:              4.52e-12       Length:        600
Score:                  310.50         Matches:       73
Percent Similarity:     58.39%         Conservative:  7
Best Local Similarity:  53.28%         Mismatches:    35
Query Match:            27.55%         Indels:        23
DB:                     12             Gaps:          3

US-09-595-947E-10 (1-214) x BG808248 (1-600)

Qy         91 ArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeu 110
              ||||||||||||||||||||:::||||||:::|||||||||||||||||||   ||||||
Db          1 CGCGAGCGCAACCGCATGCACAACCTAAACGCCCGCGCTGGACGCGCTGCGCGAGGTGCTG 60

Qy        111 ProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHisAsn 130
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATCGAGACGCTGCGCTTCGCCCACAAT 120

Qy        131 TyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeuTyrAla------ 148
              ||||||||||||||||||:::|||||||||:::|||||||||          |||
Db        121 TACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCGGACCACTGCGCCGGCGCCGGTGGC 180

Qy        149 ----------------------------LeuGluProProAlaProHisCysGlyGlu 158
                                         |||      |||              |||
Db        181 CTCCAGGGGGCGCTCTTCACGGAGGCGGTGCTCCTGAGCCCGGGAGCTGCGCTCGGCGCC 240

Qy        159 LeuGlySerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGln 178
              |||      ||||||    |||          ||||||   |||
Db        241 AGCGGGGACAGCCCTTCTCCACCTTCCTCCTGGAGCTGCACCAACAGCCCGGCGTCATCC 300

Qy        179 AlaGlySerLeuSerPro-----------------AlaAlaSerLeuGluGluArgPro 192
              :::   |||   ||||||                ||||||        |||
Db        301 TCCAACTCCACGTCCCCATACAGCTGCACTTTATCGCCCGCTAGCCC-CGGGTCAGACGT 359

Qy        193 GlyLeuLeuGlyAlaThrSerSerAla-----------CysLeuSerPro 205
              |||||||||   |||   :::|||              |||   ||||||
Db        360 GGACTACTGGCAGCCCCCACCTCCGGAGAAGCATCGTTATGCGCCTCACCT 410

RESULT 12
BU924937
LOCUS       BU924937                600 bp    mRNA    linear   EST 30-OCT-2002
DEFINITION  7103-91 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
            mRNA sequence.
ACCESSION   BU924937
VERSION     BU924937.1  GI:24428820
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
            White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
  TITLE     Gene expression in the developing mouse retina by EST sequencing
            and microarray analysis
  JOURNAL   Nucleic Acids Res. 29 (24), 4983-4993 (2001)
  MEDLINE   21671825
    PUBMED  11812828
COMMENT     Contact: Klein WH
            Department of Biochemistry and Molecular Biology
            University of Texas M.D. Anderson Cancer Center
            Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
            Tel: 713 792 3646
            Fax: 713 790 0329.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="neural retina"
                     /dev_stage="embryonic day 14.5 post-fertilization"
                     /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
BASE COUNT       99 a    194 c    222 g     82 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:              2.77e-11       Length:        600
Score:                  299.00         Matches:       88
Percent Similarity:     52.15%         Conservative:  21
Best Local Similarity:  42.11%         Mismatches:    32
Query Match:            26.53%         Indels:        68
DB:                     13             Gaps:          10

US-09-595-947E-10 (1-214) x BU924937 (1-600)

Qy         26 GluAspGluVal-----------------ThrCysProThrSerAlaProProSerPro 39
              |||:::||||||                ||   ||||||      :::|||
Db         39 GAAGAGGAGGTACTGATGCTGCTGGGCTCGGCTTCCCCGGCCTCGGCGACCCTGACCCCG 98

Qy         40 ---------------------------------ThrArgThrProGlyAsnCys--- 46
                                               |||   ||||||:::
Db         99 ATGTCCTCCAGCGCGGACGAGGAGGAGGACGAGGAGCTGCGCCGGCCGGGCTCCGCGCGT 158

Qy         47 ------------AlaGluAlaGluGlu-----------------------------Gly 52
                         |||||||||   :::                            |||
Db        159 GGGCAGCGTGGAGCGGAAGCCGGGCAGGGGGTGCAGGGCAGTCCGGCGTCGGGTGCCGGG 218

Qy         53 GlyCysArgGlyAlaProArgLysLeuArgAla-----------------ArgArgGly 66
              |||||||||      |||   :::|||                      ||||||
Db        219 GGTTGCCGG------CCAGGGCGGCTGCTGGGCCTGATGCACGAGTGCAAGCGTCGCCCG 272

Qy         67 GlyArgSerArg--------------ProLysSerGluLeuAlaLeuSerLysGlnArg 81
              |||||||||                |||:::          :::   :::   :::
Db        273 TCGCGCTCACGGGCCGTCTCCCGAGGTGCCAAGACGGCGGAGACGGTGCAGCGCATCAAG 332

Qy         82 ArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101
              ::::::||||||   |||||||||:::||||||||||||||||||:::||||||:::
Db        333 AAGACCCGCAGGCTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCC 392

Qy        102 AlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLys 121
              ||||||||||||||||||   ||||||||||||||||||:::||||||||||||||||||
Db        393 GCGCTGGACGCGCTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAG 452

Qy        122 IleGluThrLeu--ArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgI 141
              ||||||||||     ||||||||||||||||||||||||||||:::|||||||||:
Db        453 ATCGAGACGCTTGCGCTTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCC 512

Qy        141 leAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlyS 161
              ::||||||                                |||||
Db        513 TGGCGGAC--------------------------------CACTGC------------- 526

Qy        161 erProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGlyS 181
                     |||      |||      |||:::|||:::         :::   |||
Db        527 -----GCCGGCGCCGGTGGCCTNCAGGGGGCGCTCTTC---------ACGGANGCGGTGC 572

Qy        181 erLeuSerProAlaAlaSerLeu 188
              |||||||||   |||:::|||
Db        573 TNCTGAGCCCGGGAGCTGCGCTC 595

RESULT 13
AL540071
LOCUS       AL540071                1001 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL540071 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF035YD19 5-PRIME, mRNA sequence.
ACCESSION   AL540071
VERSION     AL540071.2  GI:31264632
```

```
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1001)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12869886.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            2626.r, Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF035CB10QP1.
FEATURES             Location/Qualifiers
     source          1. .1001
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF035YD19"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
BASE COUNT      190 a    322 c    342 g    121 t     26 others
ORIGIN

Alignment Scores:
Pred. No.:              4.9e-11        Length:         1001
Score:                  298.50         Matches:        81
Percent Similarity:     48.50%         Conservative:   16
Best Local Similarity:  40.50%         Mismatches:     40
Query Match:            26.49%         Indels:         64
DB:                     9              Gaps:           8

US-09-595-947E-10 (1-214) x AL540071 (1-1001)

Qy          31 CysProThrSerAlaProProSerProThrArgThrProGlyAsnCys---------- 46
               |||   :::      ||||||   |||   |||            |||
Db         421 TGCTGCTCGGATCGGCCTCCCCCGCCTTGGCGGCCCTGACCCCGCTGTCATCCAGCGCCG 480

Qy          47 ------------------AlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
                              :::   |||   :::||||||   |||         |||
Db         481 ACGAAGAAGAGGAGGAGGAGCCGGGCGCGTCAGGGCGGGGCGCGTCGGCAGGCGCGGGGC 540

Qy          61 LeuArgAlaArgArgGlyGlyArgSer-------------------------------- 69
               |||      |||      ||||||
Db         541 TTA---NGCCGGGCAANGGGGCGCGGGGCGGCGTTGCTGCGGGGTGCGGAGGGCTGCCGG 597

Qy          70 ------------------------------------ArgProLysSerGluLeuAla 76
                                                   ||||||          |||
Db         598 CCCGCACGGCTGCTGGGTCTGGTACACGATTGCAAACGGCGCCCTTCCCGGGCGCGGGCC 657

Qy          77 LeuSerLysGln------------------------------ArgArgSerArgArgLys 86
               :::|||:::                               :::::::::||||||
Db         658 GTCTCCCGAGGCGCCAAGACGGCCGAGACGGTGCAGCGCATCAAGAAGACCCGTAGACTG 717

Qy          87 LysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeu 106
               |||||||||:::|||||||||||||||||||:::||||||:::|||||||||||||||||
Db         718 AAGGCCAACAACCGCGAGCGAAACCGCATGCACAACCTCAACGCGGCACTGGACGCGCTG 777

Qy         107 ArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeuArg 126
               |||   |||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db         778 CGCGAGGTGCTCCCCACGTTCCCCGAGGACGCCAAGCTCACCAAGATCGAGACMCTGCGC 837

Qy         127 PheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu 146
               ||||||||||||||||||||   ||||||:::||||||||:::|||
Db         838 TTCGCCCACAACTACATCTGGGGACTCACSGAGACCCTGCGCCTGGCGGRT--------- 888

Qy         147 TyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyProPro 166
                                        ||||||      |||   ||||||   |||
Db         889 ------------------------MAYTGCGGGGG-GCGGCGTCG---GGGGGCATGCCG 920

Qy         167 GlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGlySerLeuSerProAlaAla 186
               |||      |||      :::           :::|||      |||   |||   |||
Db         921 GGG------GGGCTCTTCTTC-----------GAGGCMRTGTTKYTAGMSCCGGGAGCG 962

RESULT 14
BE780690
LOCUS       BE780690               1039 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601469349F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3872172 5',
            mRNA sequence.
ACCESSION   BE780690
VERSION     BE780690.1  GI:10201888
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1039)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9626  row: b  column: 13
            High quality sequence stop: 692.
FEATURES             Location/Qualifiers
     source          1. .1039
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3872172"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_67"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      228 a    351 c    337 g    122 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              6.96e-11       Length:         1039
Score:                  296.50         Matches:        95
Percent Similarity:     49.33%         Conservative:   16
Best Local Similarity:  42.22%         Mismatches:     66
Query Match:            26.31%         Indels:         53
DB:                     10             Gaps:           6

US-09-595-947E-10 (1-214) x BE780690 (1-1039)

Qy           4 GlnProSer---------GlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
               ||||||:::           :::||||||                          :::
Db         244 CAGCCCGCCTTGAGACCTGCATCTCCGACCTCGACTGCGCCAGCAGCAGCGGCAGTGACC 303

Qy          21 PheProArgAlaSerGluAspGluValThrCys--------------------ProThr 33
               :::|||   :::|||   ||||||      |||
Db         304 TATCCGGCTTCCTCACC-GACGAGGAAGACTGTGCCAGACTCCAACAGGCAGCCTCCGCT 362
```

```
Qy         34 SerAlaProProSerProThrArgThrProGlyAsnCysAlaGluAlaGluGluGlyGly 53
              |||   ||||||:::|||   |||
Db        363 TCGGGGCCGCCCGCGCCGGCCCGC---------------------------------- 386

Qy         54 CysArgGlyAlaProArgLysLeuArgAla----------------------------- 63
                ||||||||||||         ||||||
Db        387 ---AGGGGCGCGCCCAATATCTCCCGGGCGTCTGAGGTTCCAGGGGCACAGGACGACGAG 443

Qy         64 ------------ArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79
                        ||||||  ||||||:::|||   :::||||||       |||
Db        444 CAGGAGAGGCGGCGGCGCCGCGGCCGGACGCGGGTCCGCTCCGAGGCGCTGCTGCACTCG 503

Qy         80 GlnArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99
              ||||||||||||||||   ||||||||||||||||||||||||||||||:::|||
Db        504 CTGCGCAGGAGCCGGCGCGTCAAGGCCAACGATCGCGAGCGCAACCGCATGCACAACTTG 563

Qy        100 AsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeu 119
              |||:::|||||||||||||||||   |||||||||   |||||||||:::   ||||||
Db        564 AACGCGGCCCTGGACGCACTGCGCAGCGTGCTGCCTCG-TTCCCCGACGAAACCAAGCTC 622

Qy        120 ThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeu 139
              ||||||||||||||||||   |||   :::   ||||||||||||||||   :::||||||
Db        623 ACCAAAATCGAGACGCTGGC-TTCGGCTACAA-TACATCTGGGCTCTGGCCGAGACACTG 680

Qy        140 ArgIleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeu 159
              |||:::||||||       |||             |||   :::      |||   |||
Db        681 CGCCTGGCGGATCAAGGGCTGCCCGGAGGGGTGCCCGGAAGCGGCTCCTGCCGCGAAG-- 738

Qy        160 GlySerProGlyGlyProPro--------------------GlyAspTrpGlySerLeu 172
              |||   ||||||  |||                       ||||||||||
Db        739 GGCGTACCTGGCTGGCCGGACCCCAAGGCCGGCAAAAGAAGGGGAGACTGGGGCAAGGCA 798

Qy        173 TyrSerProValSer------------GlnAlaGlySerLeuSerProAlaAlaSerLeu 188
                  |||   :::              |||                 |||
Db        799 GCGGGGCCCCCGACATGAGCCCACGATGCCGCGGCTCCGAAGAAAACTAAGGCCGGGCAC 858

Qy        189 GluGluArgProGly 193
                 |||||||||
Db        859 CGGACACGACCAGGC 873
```

RESULT 15
BU612495

```
LOCUS       BU612495                730 bp    mRNA    linear   EST 20-FEB-2003
DEFINITION  UI-M-FR0-cbc-k-21-0-UI.r1 NIH_BMAP_FR0 Mus musculus cDNA clone
            UI-M-FR0-cbc-k-21-0-UI 5', mRNA sequence.
ACCESSION   BU612495
VERSION     BU612495.1  GI:23278710
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 730)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .730
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="UI-M-FR0-cbc-k-21-0-UI"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FR0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      131 a    244 c    264 g     91 t
ORIGIN
```

Alignment Scores:

Pred. No.:	6.28e-11	Length:	730
Score:	295.00	Matches:	78
Percent Similarity:	55.06%	Conservative:	20
Best Local Similarity:	43.82%	Mismatches:	28
Query Match:	26.18%	Indels:	52
DB:	13	Gaps:	8

US-09-595-947E-10 (1-214) x BU612495 (1-730)

```
Qy         16 GluThrGluArgSerPheProArgAla------------------SerGluAspGluVal 29
              |||      |||   |||   ::::::                  |||:::||||||
Db        198 GAGCCGCGTAGGATGTTCGTCAAATCTGAGACTCTGGAGTTGAAGGAGGAAGAGGAGGTA 257

Qy         30 ------------------ThrCysProThrSerAlaProProSerPro------------ 39
                                   |||   ||||||      :::|||
Db        258 CTGATGCTGCTGGGCTCGGCTTCCCCGGCCTCGGCGACCCTGACCCCGATGTCCTCCAGC 317

Qy         40 ------------------------ThrArgThrProGlyAsnCys--------------- 46
                                        |||   ||||||:::
Db        318 GCGGACGAGGAGGAGGACGAGGAGCTGCGCCGGCCGGGCTCCGCGCGTGGGCAGCGTGGA 377

Qy         47 AlaGluAlaGluGlu-----------------------------GlyGlyCysArgGly 56
              |||||||||   :::                            ||||||||||||
Db        378 GCGGAAGCCGGGCAGGGGGTGCAGGGCAGTCCGGCGTCGGGTGCCGGGGGTTGCCGG--- 434

Qy         57 AlaProArgLysLeuArgAla------------------ArgArgGlyGlyArgSerArg 70
                 |||   :::|||                       ||||||      |||||||||
Db        435 ---CCAGGGCGGCTGCTGGGCCTGATGCACGAGTGCAAGCGTCGCCCGTCGCGCTCACGG 491

Qy         71 ---------------ProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArg 85
                           |||:::      :::   :::   :::::::::||||||
Db        492 GCCGTCTCCCGAGGTGCCAAGACGGCGGAGACGGTGCAGCGCATCAAGAAGACCCGCAGG 551

Qy         86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
              |||||||||:::||||||||||||||||||||:::||||||:::||||||||||||
Db        552 CTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCGCTGGACGCG 611

Qy        106 LeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeu 125
              ||||||   ||||||||||||||||:::||||||||||||||||||||||||||||||
Db        612 CTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATCGAGACGCTG 671

Qy        126 ArgPheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArgIleAlaAsp 143
              |||||||||   :::||||||||||||||||:::||||||||:::||||||
Db        672 CGCTTCGCCCCACATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCGGAC 725
```

Search completed: February 2, 2004, 21:35:43
Job time : 2054 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 26, 2004, 18:43:11 ; Search time 5442 Seconds
(without alignments)
10975.387 Million cell updates/sec

Title: US-09-595-947E-1
Perfect score: 1460
Sequence: 1 gcaggtagcgagaggagcag.....agagtgacctaatccagtgt 1460

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1458.4	99.9	1491	6	A91167	A91167 Sequence 1
2	1458.4	99.9	1491	6	BD023626	BD023626 Polypepti
3	1458.4	99.9	1491	10	RNRELAXT	Y10619 R.norvegicu
4	1407	96.4	258815	2	AC127817	AC127817 Rattus no
5	992.2	68.0	215050	2	AC127417	AC127417 Mus muscu
c 6	976.6	66.9	138070	2	AC109783	AC109783 Mus muscu
7	900.6	61.7	1861	10	AF364300	AF364300 Mus muscu
8	819.8	56.2	5567	10	MMMATH4B	Y09167 M.musculus
9	682.4	46.7	861	6	AX698801	AX698801 Sequence
10	682.4	46.7	861	10	MMU76208	U76208 Mus musculu
11	517.4	35.4	5340	9	AF234829	AF234829 Homo sapi
c 12	515.8	35.3	165110	9	AL450311	AL450311 Human DNA
c 13	512.6	35.1	173341	2	AC021954	AC021954 Homo sapi
14	456.8	31.3	1330	9	HSA133776	AJ133776 Homo sapi
c 15	210.6	14.4	170896	2	AC011010	AC011010 Homo sapi
16	147.6	10.1	1268	6	AR023709	AR023709 Sequence
17	147.6	10.1	1268	6	AR225842	AR225842 Sequence
18	147.6	10.1	1268	9	HSU63842	U63842 Human neuro
19	147.6	10.1	1675	9	BC008687	BC008687 Homo sapi
20	147.6	10.1	1717	9	BC028226	BC028226 Homo sapi
21	147.6	10.1	79970	2	AC011428	AC011428 Homo sapi
c 22	147.6	10.1	134506	9	AC005738	AC005738 Homo sapi
c 23	147.6	10.1	151044	2	AC011432	AC011432 Homo sapi
24	145.2	9.9	1527	6	AR308547	AR308547 Sequence
25	145.2	9.9	1527	10	RNU67777	U67777 Rattus norv
26	145.2	9.9	258118	2	AC112007	AC112007 Rattus no
c 27	145.2	9.9	258880	2	AC095430	AC095430 Rattus no
28	140.4	9.6	735	10	MMU67776	U67776 Mus musculu
29	140.4	9.6	738	6	AR308548	AR308548 Sequence
30	140.4	9.6	932	10	MMMATH4C	Y09166 M.musculus
31	140.4	9.6	1315	10	MMU63841	U63841 Mus musculu
32	140.4	9.6	1333	6	AR023715	AR023715 Sequence
33	140.4	9.6	1333	6	AR225848	AR225848 Sequence
34	140.4	9.6	71538	2	AC118243	AC118243 Mus muscu
35	140.4	9.6	149268	2	AC124395	AC124395 Mus muscu
36	139.8	9.6	1385	10	MMU76207	U76207 Mus musculu
37	139.8	9.6	1412	10	MMMATH4A	Y07621 M.musculus
38	139.8	9.6	1719	5	AB065284	AB065284 Cynops py
39	139.8	9.6	10393	10	AF303001	AF303001 Mus muscu
40	139.8	9.6	123855	2	AC102600	AC102600 Mus muscu
41	139.2	9.5	770	5	AF123884	AF123884 Gallus ga
42	139.2	9.5	790	5	GGA012659	AJ012659 Gallus ga
43	139.2	9.5	1880	5	AF303000	AF303000 Gallus ga
44	139.2	9.5	2370	9	BC036847	BC036847 Homo sapi
45	139.2	9.5	6123	9	AF303002	AF303002 Homo sapi

ALIGNMENTS

RESULT 1
A91167
LOCUS A91167 1491 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent WO9827206.
ACCESSION A91167
VERSION A91167.1 GI:6740202
KEYWORDS .
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1491)
AUTHORS Icard-Liepkalns,C., Mallet,J. and Corresponding,N.A.
JOURNAL Patent: WO 9827206-A 1 25-JUN-1998;

```
          ICARD LIEPKALNS CHRISTINE (FR); MALLET JACQUES (FR)
FEATURES             Location/Qualifiers
     source          1. .1491
                     /organism="Rattus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10118"
     CDS             459. .1103
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAB69414.1"
                     /db_xref="GI:6740203"
                     /translation="MAPHPLDAPTIQVSQETQQPFPGASDHEVLSSNSTPPSPTLVPR
                     DCSEAEAGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSSGDWGSIYSPVSQAGSLSPTASLEEFPGLQVPSSPSCLLPGTLVFSDFL"
BASE COUNT      307 a    487 c    413 g    284 t
ORIGIN

  Query Match             99.9%;  Score 1458.4;  DB 6;  Length 1491;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1459;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60

Qy         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120

Qy        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180

Qy        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240

Qy        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300

Qy        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360

Qy        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420

Qy        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480

Qy        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540

Qy        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600

Qy        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660

Qy        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720

Qy        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780

Qy        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840

Qy        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900

Qy        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960

Qy        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020

Qy       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080

Qy       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140

Qy       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200

Qy       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260

Qy       1261 GCTGTGCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTGTGCGTGCACAAAGGACATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320

Qy       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380

Qy       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440

Qy       1441 AGAGTGACCTAATCCAGTGT 1460
              ||||||||||||||||||||
Db       1441 AGAGTGACCTAATCCAGTGT 1460


RESULT 2
BD023626
LOCUS       BD023626               1491 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Polypeptide belonging to the family of basic helix-loop-helix
            (bHLH) family and nucleic acid sequence corresponding thereto.
ACCESSION   BD023626
VERSION     BD023626.1  GI:22564849
KEYWORDS    JP 2001510464-A/1.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1491)
  AUTHORS   Liepkalns,C.I., Mallet,J. and Ravassard,P.
  TITLE     Polypeptide belonging to the family of basic helix-loop-helix
            (bHLH) family and nucleic acid sequence corresponding thereto
  JOURNAL   Patent: JP 2001510464-A 1 31-JUL-2001;
            RHONE POULENC RORER SA
COMMENT     OS   Rattus sp. (rat)
            PN   JP 2001510464-A/1
            PD   31-JUL-2001
            PF   19-DEC-1997 JP 1998527415
```

```
           PR   19-DEC-1996 FR     96/15651
           PI   CHRISTINE ICARD LIEPKALNS,JACQUES MALLET,PHILIPPE RAVASSARD PC
             C07K14/47,A61K31/711,A61K35/76,A61K38/00,A61K48/00,A61P25/00, PC
            C12N15/09,
           PC   C12N15/00,A61K37/02
           CC   Strandedness: Single;
           CC   Topology: Linear;
           FH   Key             Location/Qualifiers.
FEATURES             Location/Qualifiers
     source          1. .1491
                     /organism="Rattus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10118"
BASE COUNT      307 a    487 c    413 g    284 t
ORIGIN

  Query Match             99.9%;  Score 1458.4;  DB 6;  Length 1491;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1459;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60

Qy         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120

Qy        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180

Qy        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240

Qy        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300

Qy        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360

Qy        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420

Qy        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480

Qy        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540

Qy        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600

Qy        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660

Qy        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720

Qy        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780

Qy        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840

Qy        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900

Qy        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960

Qy        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020

Qy       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080

Qy       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140

Qy       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200

Qy       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260

Qy       1261 GCTGTGCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTGTGCGTGCACAAAGGACATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320

Qy       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380

Qy       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440

Qy       1441 AGAGTGACCTAATCCAGTGT 1460
              ||||||||||||||||||||
Db       1441 AGAGTGACCTAATCCAGTGT 1460


RESULT 3
RNRELAXT
LOCUS       RNRELAXT                1491 bp    mRNA    linear   ROD 06-MAY-1997
DEFINITION  R.norvegicus mRNA for transcriptional regulator, Relax.
ACCESSION   Y10619
VERSION     Y10619.1  GI:2072737
KEYWORDS    Relax; transcriptional regulator.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Ravassard,P., Chatail,F., Mallet,J. and Icard-Liepkalns,C.
  TITLE     Relax, a novel rat bHLH transcriptional regulator transiently
            expressed in the ventricular proliferating zone of the developing
            central nervous system
  JOURNAL   J. Neurosci. Res. 48 (2), 146-158 (1997)
  MEDLINE   97276390
   PUBMED   9130143
REFERENCE   2  (bases 1 to 1491)
  AUTHORS   Ravassard,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,
            Hopital de la Pitie Salpetriere, 83 Bd. de l'Hopital, F-75013
```

```
          Paris, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1491
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone_lib="RT-PCR"
     CDS             459. .1103
                     /function="transcriptional regulator"
                     /codon_start=1
                     /product="Relax"
                     /protein_id="CAA71630.1"
                     /db_xref="GI:2072738"
                     /db_xref="SPTREMBL:O08718"
                     /translation="MAPHPLDAPTIQVSQETQQPFPGASDHEVLSSNSTPPSPTLVPR
                     DCSEAEAGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSSGDWGSIYSPVSQAGSLSPTASLEEFPGLQVPSSPSCLLPGTLVFSDFL"
BASE COUNT      307 a    487 c    413 g    284 t
ORIGIN

  Query Match             99.9%;  Score 1458.4;  DB 10;  Length 1491;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1459;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60

Qy         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120

Qy        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180

Qy        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240

Qy        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300

Qy        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360

Qy        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420

Qy        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480

Qy        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540

Qy        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600

Qy        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660

Qy        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720

Qy        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780

Qy        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840

Qy        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900

Qy        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960

Qy        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020

Qy       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080

Qy       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140

Qy       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200

Qy       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260

Qy       1261 GCTGTGCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTGTGCGTGCACAAAGGACATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320

Qy       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380

Qy       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440

Qy       1441 AGAGTGACCTAATCCAGTGT 1460
              ||||||||||||||||||||
Db       1441 AGAGTGACCTAATCCAGTGT 1460


RESULT 4
AC127817
LOCUS       AC127817               258815 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-259G16, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC127817
VERSION     AC127817.3  GI:25077905
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 258815)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
```

Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 258815)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 258815)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Nov 19, 2002 this sequence version replaced gi:23912578.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GZXS
Center clone name: CH230-259G16
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 224747 bases at least Q40
Consensus quality: 227981 bases at least Q30
Consensus quality: 229752 bases at least Q20
Estimated insert size: 228243; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 255979: contig of 255979 bp in length
* 255980 256079: gap of unknown length
* 256080 257349: contig of 1270 bp in length
* 257350 257449: gap of unknown length
* 257450 258815: contig of 1366 bp in length.

FEATURES Location/Qualifiers

source 1. .258815
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-259G16"

misc_feature 1. .1609
/note="wgs_end_extension
clone_end:Sp6"

misc_feature complement(2865. .3659)
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BZ169379"

misc_feature 224098. .225157
/note="clone_boundary
clone_end:T7
site:
end_sequence:BZ169378"

BASE COUNT 59351 a 56888 c 58131 g 56821 t 27624 others

ORIGIN

Query Match 96.4%; Score 1407; DB 2; Length 258815;
Best Local Similarity 99.1%; Pred. No. 2.5e-305;
Matches 1446; Conservative 0; Mismatches 10; Indels 3; Gaps 3;

```
Qy          4 GGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCA 63
              ||   || | || ||||||||||||||||||||||||||||||||||||||||||||||
Db     247404 GGGCTCGCGTGGCGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCA 247463

Qy         64 GCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGG-ACACACG 122
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     247464 GCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGCACACACG 247523

Qy        123 ATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCTCC 182
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     247524 ATTAGCAGCTCAGAAATCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCTCC 247583

Qy        183 GAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAGCA 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247584 GAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAGCA 247643

Qy        243 AAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACC-ACCCG 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
```

```
Db     247644 AAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCAACCCG 247703

Qy        302 CAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGGT 361
              ||||||||||||||||||||||||| |||||||||||||||||  ||||||||||||
Db     247704 CAGCTCTCTGTTCTTTTGAGCCCGGAG-AACTAGGTAACATTTAGGCCCCTCCAAAGGGT 247762

Qy        362 AGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGA 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247763 AGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGA 247822

Qy        422 CTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGC 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247823 CTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGC 247882

Qy        482 GCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGA 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247883 GCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGA 247942

Qy        542 AGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGA 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247943 AGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGA 248002

Qy        602 AGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCG 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248003 AGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCG 248062

Qy        662 CAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGC 721
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248063 CAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGC 248122

Qy        722 CAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGG 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248123 CAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGG 248182

Qy        782 TGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGC 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248183 TGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGC 248242

Qy        842 CCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGG 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248243 CCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGG 248302

Qy        902 CCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGA 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248303 CCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGA 248362

Qy        962 CTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATT 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248363 CTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATT 248422

Qy       1022 GGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCT 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248423 GGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCT 248482

Qy       1082 GGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAGA 1141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248483 GGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAGA 248542

Qy       1142 AAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGCC 1201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248543 AAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGCC 248602

Qy       1202 CCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCTG 1261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248603 CCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCTG 248662

Qy       1262 CTGTGCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCCA 1321
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db     248663 CTGTGCGTGCACAAAGGACATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCCA 248722

Qy       1322 CCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCCA 1381
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248723 CCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCCA 248782

Qy       1382 TACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCCA 1441
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248783 TACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCCA 248842

Qy       1442 GAGTGACCTAATCCAGTGT 1460
              |||||||||||||||||||
Db     248843 GAGTGACCTAATCCAGTGT 248861
```

```
RESULT 5
AC127417
LOCUS       AC127417               215050 bp    DNA     linear   HTG 19-OCT-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-459M2, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC127417
VERSION     AC127417.2  GI:24137619
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 215050)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 215050)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 215050)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Oct 19, 2002 this sequence version replaced gi:21759524.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: M_BA0459M02
            ----------
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 212279 bases at least Q40
            Consensus quality: 213137 bases at least Q30
            Consensus quality: 213590 bases at least Q20
            Insert size: 192000; agarose-fp
            Insert size: 217944; sum-of-contigs
            Quality coverage: 11.89 in Q20 bases; agarose-fp
            Quality coverage: 10.54 in Q20 bases; sum-of-contigs
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4227: contig of 4227 bp in length
            *     4228     4327: gap of unknown length
            *     4328    16589: contig of 12262 bp in length
            *    16590    16689: gap of unknown length
            *    16690    27318: contig of 10629 bp in length
```

```
            *    27319    27418: gap of unknown length
            *    27419    49797: contig of 22379 bp in length
            *    49798    49897: gap of unknown length
            *    49898    79378: contig of 29481 bp in length
            *    79379    79478: gap of unknown length
            *    79479   132595: contig of 53117 bp in length
            *   132596   132695: gap of unknown length
            *   132696   215050: contig of 82355 bp in length.
FEATURES             Location/Qualifiers
     source          1. .215050
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-459M2"
     misc_feature    1. .4227
                     /note="assembly_name:Contig31"
     misc_feature    4328. .16589
                     /note="assembly_name:Contig32"
     misc_feature    16690. .27318
                     /note="assembly_name:Contig33"
     misc_feature    27419. .49797
                     /note="assembly_name:Contig34"
     misc_feature    49898. .79378
                     /note="assembly_name:Contig35"
     misc_feature    79479. .132595
                     /note="assembly_name:Contig36"
     misc_feature    132696. .215050
                     /note="assembly_name:Contig37"
BASE COUNT    56024 a  52513 c  52885 g  53012 t    616 others
ORIGIN

  Query Match             68.0%;  Score 992.2;  DB 2;  Length 215050;
  Best Local Similarity   83.9%;  Pred. No. 2.8e-212;
  Matches 1275;  Conservative    0;  Mismatches  178;  Indels   66;  Gaps   11;

Qy          4 GGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCA 63
              ||   || | || |||| ||||||||||||   |||||||||||||||  |||||||||
Db     205068 GGGCTCGCGTGGCGCAGCCCCTGGGCCCCCTCCGCTGATTGGCCCGTGGTGCAGGCAGCA 205127

Qy         64 GCCCGGCAGGCACGCTCCTGGTC-CGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 122
              |||||||||||||||||||| |  ||||||||||||||||||||||||||||||||||
Db     205128 GCCCGGCAGGCACGCTCCTGGCCGGGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 205187

Qy        123 ATT----AGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCA-CAGAGGCCGAGGACCC 177
              | |      |||||||||| |||||||||||||| ||||||| ||| || |||| |||
Db     205188 ACTTGCATGCAGCTCAGAAATCCCTCTGGGTCTCATCACTGCAGCAGTGGTCGAGTACCT 205247

Qy        178 CCTCCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCT 237
              |||| |||||| | | |  | ||||||||||||||||||||||||||||||||||| |
Db     205248 CCTCGGAGCTTTTCTACGACTTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGTT 205307

Qy        238 CAGCAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACC- 296
              |||| ||||||  || |||||||| || |||||||||| |||||||||| ||||||
Db     205308 TAGCAGAACTTCAGAGGGAGCAGAGAGGCTCAGCTATCCACTGCTGCTTGACACTGACCC 205367

Qy        297 --------------------------ACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTA 330
                                       ||| || ||||||| |||||||||| | |||| |
Db     205368 TATCCACTGCTGCTTGTCACTGACTGACCTGCTGCTCTCTATTCTTTTGAGTCGGGAG-A 205426

Qy        331 ACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGGAGTG----------GGTGG 380
              |||||||||| || |||| |||||||||| || |||||| | |           | ||
Db     205427 ACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGGCGCGCGGGGTGTGTGTGGG 205486

Qy        381 GCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTGTCACACCCC--CCTTC 438
              |  |||||| ||||| |||| |||||||||||||||||| ||| ||||| |   |||||
Db     205487 GGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGGCTGGCACACACACACCTTC 205546

Qy        439 CATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGT 498
              |||||||||||||| |||||||||||||||||||||||||||| ||||||||||||||
Db     205547 CATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGT 205606

Qy        499 CCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 558
              ||| |||||| || ||||| |||||||||||||||||||||||||||||||||||||||
Db     205607 CCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 205666

Qy        559 CCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACT 618
              |||||||||||||||||||||| |||| ||||||||||||||||||||  |||||||
Db     205667 CCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACT 205726

Qy        619 GCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCG 678
              |||||||||| |||||||||||||| || || |||||||||||||||||||||||||||
Db     205727 GCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCG 205786

Qy        679 AGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCA 738
              ||||||||| ||||| |||||| |||||||||||||||||||| || ||||||||||
Db     205787 AGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCA 205846

Qy        739 ACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 798
              | |||||||||||||| ||||| |||||||||||||||||||||||||||||||||||
Db     205847 ATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 205906

Qy        799 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGG 858
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     205907 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGG 205966

Qy        859 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGC 918
              |||||||||||||||||||||||||||||||||||| ||||| ||||||||||||||
Db     205967 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGC 206026

Qy        919 CCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACT 978
              ||||||| |||||||| ||||| ||||| ||||||| ||| ||||||||||||||||||
Db     206027 CCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACT 206086

Qy        979 CCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCC 1038
              ||||||| |||||||| |||| |||||||||||| |||||||||||||| ||||||||||
Db     206087 CCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCC 206146

Qy       1039 TGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCT 1098
              |||||||||||||||||||||| ||||||||||  | ||||||||||||||||||||
Db     206147 TGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCT 206206

Qy       1099 TGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAG 1158
              ||||||| | ||   | ||| ||||| |||||| ||| |  ||||||||||   |||||
Db     206207 TGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTGGAAAGGGAGGGGACCAGAG 206266

Qy       1159 CTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGCCCCTTCTGGCTTTCATTA 1218
              | ||||| | ||| |||||||||||||| ||| |||||||||| |||||||||||| ||
Db     206267 CCGTCTGGAGTGGGAGGTAGTGGAGGCTCTCAAGCATCTCGCCTCTTCTGGCTTTCACTA 206326

Qy       1219 GTCAGGTCCCT------------GATTTAACCAGGATTCGCACAGTTCCTTGCTGCTGT 1265
              | | |||||              |  ||||| ||| ||| ||  | | |   ||
Db     206327 CTTGGATCCCTAGCCCTCTCACAGGGCTTAACTAGGCTTCTCATCGGTAC--CCTTGCTG 206384

Qy       1266 GCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCCACCTC 1325
              | |||||   || |||| |||||| ||| |||||||| |||||| ||| |||||  ||
Db     206385 CTGCGCACAGCAGACATTGGGGGCTGCTCTTCTCTTAACTCTCCTCGGTGCAGCCACATC 206444

Qy       1326 AAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCCATACT 1385
              |||||| |||||||||||   ||||   |||||||||  ||||| |||||||||||||||
Db     206445 AAACTCTCGCTCCAAGCATTTGAGAATGGTAGCACTACCTAGTT-GGAGACTCCCATACT 206503

Qy       1386 TCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCT----CCAACCACCGCTTTCTCCA 1441
              ||||||||| || |||||| |||||||||| |||||    ||| ||| | |||| ||||
Db     206504 TCCTGGTGAGTCTGCCCTCATTCAAATCTGCCGGCCTCCGACCATCCATCACTTTTTCCA 206563

Qy       1442 GAGTGACCTAATCCAGTGT 1460
              | ||||||||||||||||||
Db     206564 GGGTGACCTAATCCAGTGT 206582


RESULT 6
AC109783/c
LOCUS       AC109783              138070 bp    DNA     linear   HTG 07-FEB-2002
DEFINITION  Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
            unordered pieces.
```

```
ACCESSION   AC109783
VERSION     AC109783.1  GI:18581594
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 138070)
  AUTHORS   McCombie,W.R., Baker,J.P., Balija,V., Dedhia,N.N., de la
            Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
            Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,
            Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
            Zutavern,T.
  TITLE     Mouse Genomic Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 138070)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
COMMENT     -------------- Genome Center --------------
            Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
            Laboratory
            Center code: CSHL
            Web site: http://www.cshl.org/genseq
            Contact:  mccombie@cshl.org
            ------------------- Project Information
            Project name: RP23-121F10
            Clone name: RP23-121F10
            Insert size: 173000; agarose-fp
            Insert size: 141616; sum-of-contigs
            Quality coverage: 4.00 in Q20 bases; agarose-fp
            Quality coverage: 3.70 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    17785: contig of 17785 bp in length
            *   17786    17874: gap of unknown length
            *   17875    30356: contig of 12482 bp in length
            *   30357    30444: gap of unknown length
            *   30445    42306: contig of 11862 bp in length
            *   42307    42394: gap of unknown length
            *   42395    53598: contig of 11204 bp in length
            *   53599    53686: gap of unknown length
            *   53687    64355: contig of 10669 bp in length
            *   64356    64443: gap of unknown length
            *   64444    74016: contig of 9573 bp in length
            *   74017    74104: gap of unknown length
            *   74105    83366: contig of 9262 bp in length
            *   83367    83454: gap of unknown length
            *   83455    92355: contig of 8901 bp in length
            *   92356    92443: gap of unknown length
            *   92444   100821: contig of 8378 bp in length
            *  100822   100909: gap of unknown length
            *  100910   107529: contig of 6620 bp in length
            *  107530   107617: gap of unknown length
            *  107618   114066: contig of 6449 bp in length
            *  114067   114154: gap of unknown length
            *  114155   118873: contig of 4719 bp in length
            *  118874   118961: gap of unknown length
            *  118962   123619: contig of 4658 bp in length
            *  123620   123707: gap of unknown length
            *  123708   128240: contig of 4533 bp in length
            *  128241   128328: gap of unknown length
            *  128329   132682: contig of 4354 bp in length
            *  132683   132770: gap of unknown length
            *  132771   136341: contig of 3571 bp in length
            *  136342   136429: gap of unknown length
            *  136430   138070: contig of 1641 bp in length.
FEATURES             Location/Qualifiers
     source          1. .138070
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-121F10"
BASE COUNT    35731 a  33657 c  33303 g  33954 t   1425 others
ORIGIN
```

Query Match 66.9%; Score 976.6; DB 2; Length 138070;
Best Local Similarity 83.5%; Pred. No. 9.2e-209;
Matches 1263; Conservative 0; Mismatches 184; Indels 66; Gaps 11;

```
Qy         10 GAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGT-GGCACAGGCAGCAGCCCG 68
Db     111713 GCGTGGCCCAGCCCCTAGGGCCCCTCCGCTGATTGGCCCGTGGGTGCAGGCAGCAGCCCG 111654

Qy         69 GCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACGATT--- 125
Db     111653 CCAGGCACGCTCCTGGCGGAGCCAGAGCAGATAAAGCGTGCCAGGGAACACACGATTTGC 111594

Qy        126 -AGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCA-CAGAGGCCGAGGACCCCCTCCG 183
Db     111593 ATGCAGCTCAGAAATCCCTCTGGGTCTCATCACTGCAGCAGTGGTCGAGTACCTCCTCGG 111534

Qy        184 AGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAGCAA 243
Db     111533 AGCTTTTTTACGACTTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGTTTAGCAG 111474

Qy        244 AACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACC------- 296
Db     111473 AACTTCAGAGGGAGCAGAGAGGCTCAGCTATCCACTGCTGCTTGACACTGACCCTATCCA 111414

Qy        297 --------------------ACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGG 336
Db     111413 CTGCTGCTTGTCACTGACTGACCTGCTGCTCTCTATTCTTTTGAGTCGGGAG-AACTAGG 111355

Qy        337 TAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGGAGTG----------GGTGGGCGTAC 386
Db     111354 TAACAATTCGGAAACTCCAAAGGGTGGATGAGGGGCGCGCGGGGTGTGTGTGGGGGATAC 111295

Qy        387 TCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTGTCACACCCC--CCTTCCATTTT 444
Db     111294 TCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGGCTGGCACACACACACCTTCCATTTT 111235

Qy        445 TTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAG 504
Db     111234 TTCCCAACCGCAGGATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAG 111175

Qy        505 AGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCC 564
Db     111174 AGACACAACAACCTTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCC 111115

Qy        565 CACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAG 624
Db     111114 CACCTAGCCCCACTCTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAG 111055

Qy        625 GGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGG 684
Db     111054 GGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGG 110995

Qy        685 CACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCA 744
Db     110994 CACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCA 110935

Qy        745 TGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATG 804
Db     110934 TGCACAACCTCAACTCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATG 110875

Qy        805 ACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGA 864
```

```
Db    110874 ACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGA 110815

Qy       865 CTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCCTGTG 924

Db    110814 CTCAGACGCTGCGCATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCCTGTG 110755

Qy       925 GGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAG 984

Db    110754 GAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAG 110695

Qy       985 TTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGG 1044

Db    110694 TCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGG 110635

Qy      1045 TGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTGTGAA 1104

Db    110634 TGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTGTGAA 110575

Qy      1105 GGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAGCTGTCT 1164

Db    110574 GAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTGGAAAGGGAGGGGACCACCACCGTCT 110515

Qy      1165 GAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGCCCCTTCTGGCTTTCATTAGTCAGG 1224

Db    110514 GGAGTGGGAGGTAGTGGAGGCTCTCAAGCATCTCGCCTCTTCTGGCTTTCACTACTTGGA 110455

Qy      1225 TCCCT------------GATTTAACCAGGATTCGCACAGTTCCTTGCTGCTGTGCGTGC 1271

Db    110454 TCCCTAGCCCTCTCACAGGGCTTAACTAGGCTTCTCATCGGTAC--CCTTGCTGCTGCGC 110397

Qy      1272 ACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCCACCTCAAACTC 1331

Db    110396 ACAGCAGACATTGGGGGCTGCTCTTCTCTTAACTCTCCTCGGTGCAGCCACATCAAACTC 110337

Qy      1332 CCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCCATACTTCCTGG 1391

Db    110336 TCGCTCCAAGCATTTGAGAATGGTAGCACTACCTAGTT-GGAGACTCCCATACTTCCTGG 110278

Qy      1392 TGACTCCGCCCTCTTTCAAATCTGCGGGCCT----CCAACCACCGCTTTCTCCAGAGTGA 1447

Db    110277 TGAGTCTGCCCTCATTCAAATCTGCCGGCCTCCGACCATCCATCACTTTTTCCAGGGTGA 110218

Qy      1448 CCTAATCCAGTGT 1460

Db    110217 CCTAATCCAGTGT 110205


RESULT 7
AF364300
LOCUS       AF364300               1861 bp    DNA     linear   ROD 03-MAY-2001
DEFINITION  Mus musculus neurogenin 3 gene, complete cds.
ACCESSION   AF364300
VERSION     AF364300.1  GI:13937128
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1861)
  AUTHORS   Lee,J., Smith,S., Watada,H., Lin,J., Scheel,D., Wang,J., Mirmira,R.
            and German,M.
  TITLE     Regulation of the pancreatic pro-endocrine gene neurogenin3
  JOURNAL   Diabetes (2001) In press
REFERENCE   2  (bases 1 to 1861)
  AUTHORS   Schwitzgebel,V. and German,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2001) Hormone Research Institute, University of
            California San Francisco, 513 Parnassus Ave, HSW1090, San
            Francisco, CA 94145-0534, USA
FEATURES             Location/Qualifiers
     source          1. .1861
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     mRNA            <1093. .>1737
                     /product="neurogenin 3"
     CDS             1093. .1737
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="AAK50058.1"
                     /db_xref="GI:13937129"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT      397 a    560 c    537 g    367 t
ORIGIN

  Query Match             61.7%;  Score 900.6;  DB 10;  Length 1861;
  Best Local Similarity   86.2%;  Pred. No. 1.7e-191;
  Matches 1094;  Conservative    0;  Mismatches  129;  Indels   46;  Gaps    7;

Qy          4 GGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCA 63

Db        594 GGGCTCGCGTGGCGCAGCCCCTGGGCCCCCTCCGCTGATTGGCCCGTGGTGCAGGCAGCA 653

Qy         64 GCCCGGCAGGCACGCTCCTGGTC-CGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 122

Db        654 GCCCGGCAGGCACGCTCCTGGCCGGGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 713

Qy        123 ATT----AGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCA-CAGAGGCCGAGGACCC 177

Db        714 ACTTGCATGCAGCTCAGAAATCCCTCTGGGTCTCATCACTGCAGCAGTGGTCGAGTACCT 773

Qy        178 CCTCCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCT 237

Db        774 CCTCGGAGCTTTTCTACGACTTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGTT 833

Qy        238 CAGCAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACC- 296

Db        834 TAGCAGAACTTCAGAGGGAGCAGAGAGGCTCAGCTATCCACTGCTGCTTGACACTGACCC 893

Qy        297 --------------------------ACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTA 330

Db        894 TATCCACTGCTGCTTGTCACTGACTGACCTGCTGCTCTCTATTCTTTTGAGTCGGGAG-A 952

Qy        331 ACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGGAGTGGGT----------GG 380

Db        953 ACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGGCGCGCGGGGTGTGTGTGGG 1012

Qy        381 GCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTG--TCACACCCCCCTTC 438

Db       1013 GGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGGCTGGCACACACACACCTTC 1072

Qy        439 CATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGT 498

Db       1073 CATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGT 1132

Qy        499 CCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 558

Db       1133 CCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 1192

Qy        559 CCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACT 618

Db       1193 CCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACT 1252

Qy        619 GCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCG 678

Db       1253 GCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCG 1312

Qy        679 AGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCA 738

Db       1313 AGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCA 1372

Qy        739 ACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 798
```

```
Db       1373 ATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 1432

Qy        799 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGG 858

Db       1433 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGG 1492

Qy        859 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGC 918

Db       1493 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGC 1552

Qy        919 CCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACT 978

Db       1553 CCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACT 1612

Qy        979 CCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCC 1038

Db       1613 CCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCC 1672

Qy       1039 TGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCT 1098

Db       1673 TGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCT 1732

Qy       1099 TGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAG 1158

Db       1733 TGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTGGAAAGGGAGGGGACCAGAG 1792

Qy       1159 CTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGCCCCTTCTGGCTTTCATTA 1218

Db       1793 CCGTCTGGAGTGGGAGGTAGTGGAGGCTCTCAAGCATCTCGCCTCTTCTGGCTTTCACTA 1852

Qy       1219 GTCAGGTCC 1227

Db       1853 CTTGGATCC 1861
```

```
RESULT 8
MMMATH4B
LOCUS       MMMATH4B                5567 bp    DNA     linear   ROD 25-OCT-2000
DEFINITION  M.musculus MATH4B gene.
ACCESSION   Y09167
VERSION     Y09167.2  GI:11065737
KEYWORDS    bHLH; mammalian atonal homolog 4B; MATH4B; neurogenin 3.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Cau,E., Gradwohl,G., Fode,C. and Guillemot,F.
  TITLE     Mash1 activates a cascade of bHLH regulators in olfactory neuron
            progenitors
  JOURNAL   Development 124 (8), 1611-1621 (1997)
  MEDLINE   97261963
   PUBMED   9108377
REFERENCE   2
  AUTHORS   Jacquemin,P., Durviaux,S.M., Jensen,J., Godfraind,C., Gradwohl,G.,
            Guillemot,F., Madsen,O.D., Carmeliet,P., Dewerchin,M., Collen,D.,
            Rousseau,G.G. and Lemaigre,F.P.
  TITLE     Transcription factor hepatocyte nuclear factor 6 regulates
            pancreatic endocrine cell differentiation and controls expression
            of the proendocrine gene ngn3
  JOURNAL   Mol. Cell. Biol. 20 (12), 4445-4454 (2000)
  MEDLINE   20285449
   PUBMED   10825208
REFERENCE   3
  AUTHORS   Gradwohl,G.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC,
            CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg,
            F-67404 ILLKIRCH cedex, FRANCE
  REMARK    Revised by [4]
REFERENCE   4  (bases 1 to 5567)
  AUTHORS   Lemaigre,F.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic
            Research Unit, Louvain University Medical School, Avenue Hippocrate
            75, box 7529, Brussels 1200, BELGIUM
COMMENT     On Oct 31, 2000 this sequence version replaced gi:1666087.
FEATURES             Location/Qualifiers
     source          1. .5567
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="SV129"
                     /db_xref="taxon:10090"
                     /clone="lambda GEM12 clone 19C"
                     /cell_line="D3"
                     /cell_type="embryonic stem cells"
                     /clone_lib="Genomic library No.62, MboI partial digests
                     cloned in BamHI"
                     /dev_stage="embryos"
     misc_binding    958
                     /bound_moiety="HNF6"
     misc_binding    3692
                     /bound_moiety="HNF6"
     gene            4145. .5567
                     /gene="MATH4B"
     TATA_signal     4145. .4151
                     /gene="MATH4B"
                     /note="putative"
     CDS             4923. .5567
                     /gene="MATH4B"
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="CAA70366.1"
                     /db_xref="GI:1666088"
                     /db_xref="SWISS-PROT:P70661"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT     1271 a   1549 c   1564 g   1183 t
ORIGIN
```

```
  Query Match           56.2%;  Score 819.8;  DB 10;  Length 5567;
  Best Local Similarity  87.1%;  Pred. No. 2e-173;
  Matches  997;  Conservative    0;  Mismatches  102;  Indels   46;  Gaps    7;

Qy           4 GGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCA 63

Db        4424 GGGCTCGCGTGGCGCAGCCCCTGGGCCCCCTCCGCTGATTGGCCCGTGGTGCAGGCAGCA 4483

Qy          64 GCCCGGCAGGCACGCTCCTGGTC-CGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 122

Db        4484 GCCCGGCAGGCACGCTCCTGGCCGGGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 4543

Qy         123 ATT----AGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCA-CAGAGGCCGAGGACCC 177

Db        4544 ACTTGCATGCAGCTCAGAAATCCCTCTGGGTCTCATCACTGCAGCAGTGGTCGAGTACCT 4603

Qy         178 CCTCCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCT 237

Db        4604 CCTCGGAGCTTTTCTACGACTTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGTT 4663

Qy         238 CAGCAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACC- 296

Db        4664 TAGCAGAACTTCAGAGGGAGCAGAGAGGCTCAGCTATCCACTGCTGCTTGACACTGACCC 4723

Qy         297 -------------------------ACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTA 330

Db        4724 TATCCACTGCTGCTTGTCACTGACTGACCTGCTGCTCTCTATTCTTTTGAGTCGGGAG-A 4782

Qy         331 ACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGGAGTG----------GGTGG 380

Db        4783 ACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGGCGCGCGGGGTGTGTGTGGG 4842

Qy         381 GCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTGTCACACCCC--CCTTC 438
```

```
Db       4843 GGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGGCTGGCACACACACACCTTC 4902

Qy        439 CATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGT 498
              |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db       4903 CATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGT 4962

Qy        499 CCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 558
              ||| |||||| || ||||| ||||||||||||||||||||||||||||||||||||||||
Db       4963 CCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 5022

Qy        559 CCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACT 618
              |||||||||||||||||||||| |||| ||||||||||||||||||||  |||||||
Db       5023 CCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACT 5082

Qy        619 GCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCG 678
              |||||||||| |||||||||||||| || || ||||||||||||||||||||||||||||
Db       5083 GCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCG 5142

Qy        679 AGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCA 738
              |||||||||| ||||| |||||| |||||||||||||||||||| || ||||||||||
Db       5143 AGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCA 5202

Qy        739 ACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 798
              | |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||
Db       5203 ATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 5262

Qy        799 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGG 858
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       5263 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGG 5322

Qy        859 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGC 918
              |||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||||
Db       5323 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGC 5382

Qy        919 CCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACT 978
              ||||||| |||||||| ||||| ||||| ||||||| ||| ||||||||||||||||||
Db       5383 CCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACT 5442

Qy        979 CCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCC 1038
              ||||||| |||||||| |||| |||||||||||| |||||||||||||| ||||||||||
Db       5443 CCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCC 5502

Qy       1039 TGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCT 1098
              |||||||||||||||||||||| ||||||||||| | ||||||||||||||||||||
Db       5503 TGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCT 5562

Qy       1099 TGTGA 1103
              |||||
Db       5563 TGTGA 5567


RESULT 9
AX698801
LOCUS       AX698801                861 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 7 from Patent WO02086107.
ACCESSION   AX698801
VERSION     AX698801.1  GI:29499589
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Wobus,A.M., St-Onge,L., Blyszczuk,P. and Hoffmann,U.
  TITLE     A method for differentiating stem cells into insulin-producing
            cells
  JOURNAL   Patent: WO 02086107-A 7 31-OCT-2002;
            DeveloGen Aktiengesellschaft fuer Entwicklungsbiologische Forschung
            (DE) ; INSTITUT FUER PFLANZENGENETIK UND KULTURPFLANZENFORSCHUNG
            (DE)
FEATURES             Location/Qualifiers
     source          1. .861
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     CDS             160. .804
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD87442.1"
                     /db_xref="GI:29499590"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT      182 a    274 c    250 g    155 t
ORIGIN

  Query Match           46.7%;  Score 682.4;  DB 6;  Length 861;
  Best Local Similarity  89.7%;  Pred. No. 1.6e-142;
  Matches  771;  Conservative    0;  Mismatches   76;  Indels   13;  Gaps    3;

Qy        312 TTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGG 371
              |||||||||| | |||| |||||||||||| || |||| |||||||||| || ||||||
Db          2 TTCTTTTGAGTCGGGAG-AACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGG 60

Qy        372 AGTGGGT----------GGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGA 421
              | | |           |||  |||||| |||| |||| ||||||||||||||||||
Db         61 CGCGCGGGGTGTGTGTGGGGGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGG 120

Qy        422 CTG--TCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGAT 479
              |||    |||| | ||||||||||||||||| ||||||||||||||||||||||||||
Db        121 CTGGCACACACACACCTTCCATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGAT 180

Qy        480 GCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCAC 539
              |||| |||||||||||||||| |||||| || ||||| ||||||||||||||||||||||
Db        181 GCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCAC 240

Qy        540 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCC 599
              |||||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db        241 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCC 300

Qy        600 GAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGG 659
              |||||||||| |||||||||||||||||| |||||||||||||| || || |||||||||
Db        301 GAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGG 360

Qy        660 CGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAG 719
              |||||||||||||||||||||||||||||| ||||| |||||| ||||||||||||||||||
Db        361 CGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAG 420

Qy        720 GCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGC 779
              ||||| || |||||||||||| |||||||||||||| ||||| ||||||||||||||||||
Db        421 GCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGC 480

Qy        780 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 839
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 540

Qy        840 GCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAC 899
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAT 600

Qy        900 GGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGC 959
              ||||| |||||||||||||||||| |||||||| ||||| ||||| ||||||| |||
Db        601 GGCCCGGAGCCCCCTGTGCCCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGG 660

Qy        960 GACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCA 1019
              |||||||||||||||||||||||||||||| |||||||| |||| ||||||||||| ||||||
Db        661 GACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCA 720

Qy       1020 TTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACC 1079
              |||||||| |||||||||||||||||||||||||||||| |||||||||||  |
Db        721 TTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCA 780

Qy       1080 CTGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCA 1139
              |||||||||||||||||||||||| | ||   | ||| ||||| |||||| ||| |
```

```
Db        781 CTGGTGTTCTCAGACTTCTTGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTG 840

Qy       1140 GAAAGGGAGGGAGTCAGAGC 1159
              ||||||||||   || |||
Db        841 GAAAGGGAGGGGACCACAGC 860


RESULT 10
MMU76208
LOCUS       MMU76208                 861 bp    DNA     linear   ROD 05-FEB-1997
DEFINITION  Mus musculus neurogenin 3 (ngn3) gene, complete cds.
ACCESSION   U76208
VERSION     U76208.1  GI:1815654
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 861)
  AUTHORS   Sommer,L., Ma,Q. and Anderson,D.J.
  TITLE     neurogenins, a novel family of atonal-related bHLH transcription
            factors, are putative mammalian neuronal determination genes that
            reveal progenitor cell heterogeneity in the developing CNS and PNS
  JOURNAL   Mol. Cell. Neurosci. 8 (4), 221-241 (1996)
  MEDLINE   97153565
   PUBMED   9000438
REFERENCE   2  (bases 1 to 861)
  AUTHORS   Sommer,L., Ma,Q. and Anderson,D.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-1996) Biology 216-76, California Institute of
            Technology, Howard Hughes Medical Institute, Pasadena, CA 91125,
            USA
COMMENT     On Feb 5, 1997 this sequence version replaced gi:1666911.
FEATURES             Location/Qualifiers
     source          1. .861
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene            160. .804
                     /gene="ngn3"
     CDS             160. .804
                     /gene="ngn3"
                     /function="bHLH transcription factor"
                     /note="NGN3"
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="AAC53029.1"
                     /db_xref="GI:1666912"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT      182 a    274 c    250 g    155 t
ORIGIN

  Query Match             46.7%;  Score 682.4;  DB 10;  Length 861;
  Best Local Similarity   89.7%;  Pred. No. 1.6e-142;
  Matches  771;  Conservative    0;  Mismatches   76;  Indels   13;  Gaps    3;

Qy        312 TTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGG 371
              |||||||||| | |||| |||||||||||| || |||| |||||||||| || ||||||
Db          2 TTCTTTTGAGTCGGGAG-AACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGG 60

Qy        372 AGTGGGT----------GGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGA 421
              | | |           |||  |||||| ||||| |||| ||||||||||||||||||
Db         61 CGCGCGGGGTGTGTGTGGGGGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGG 120

Qy        422 CTG--TCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGAT 479
              |||   ||||| | ||||||||||||||||||| ||||||||||||||||||||||||
Db        121 CTGGCACACACACACCTTCCATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGAT 180

Qy        480 GCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCAC 539
              |||| |||||||||||||||| |||||| || ||||| ||||||||||||||||||||||
```

```
Db        181 GCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCAC 240

Qy        540 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCC 599
              |||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db        241 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCC 300

Qy        600 GAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGG 659
              |||||||||| |||||||||||||||| |||||||||||||| || || |||||||||
Db        301 GAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGG 360

Qy        660 CGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAG 719
              |||||||||||||||||||||||||||||| ||||| |||||| ||||||||||||||||||
Db        361 CGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAG 420

Qy        720 GCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGC 779
              ||||| || |||||||||| |||||||||||||| ||||| ||||||||||||||||||
Db        421 GCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGC 480

Qy        780 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 540

Qy        840 GCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAC 899
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAT 600

Qy        900 GGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGC 959
              ||||| |||||||||||||||||| |||||||| ||||| ||||| ||||||| |||
Db        601 GGCCCGGAGCCCCCTGTGCCCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGG 660

Qy        960 GACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCA 1019
              |||||||||||||||||||||||||||||| |||||||| |||| ||||||||||| ||||||
Db        661 GACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCA 720

Qy       1020 TTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACC 1079
              |||||||| |||||||||||||||||||||||||||||||||| |||||||||||  |
Db        721 TTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCA 780

Qy       1080 CTGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCA 1139
              |||||||||||||||||||||||| | ||   | ||| ||||| |||||| ||| |
Db        781 CTGGTGTTCTCAGACTTCTTGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTG 840

Qy       1140 GAAAGGGAGGGAGTCAGAGC 1159
              ||||||||||   || |||
Db        841 GAAAGGGAGGGGACCACAGC 860


RESULT 11
AF234829
LOCUS       AF234829                 5340 bp    DNA     linear   PRI 19-OCT-2001
DEFINITION  Homo sapiens neurogenin 3 gene, complete cds.
ACCESSION   AF234829
VERSION     AF234829.1  GI:13183002
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5340)
  AUTHORS   del Bosque-Plata,L., Lin,J., Horikawa,Y., Schwarz,P.E., Cox,N.J.,
            Iwasaki,N., Ogata,M., Iwamoto,Y., German,M.S. and Bell,G.I.
  TITLE     Mutations in the coding region of the neurogenin 3 gene (NEUROG3)
            are not a common cause of maturity-onset diabetes of the young in
            Japanese subjects
  JOURNAL   Diabetes 50 (3), 694-696 (2001)
  MEDLINE   21140923
   PUBMED   11246894
REFERENCE   2  (bases 1 to 5340)
  AUTHORS   Lin,J. and German,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2000) Hormone Research Institute, University of
            California San Francisco, 513 Parnassus Ave., San Francisco, CA
            94143-0534, USA
```

```
FEATURES             Location/Qualifiers
     source          1. .5340
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     mRNA            <3022. .>3666
                     /product="neurogenin 3"
     CDS             3022. .3666
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="AAK15022.1"
                     /db_xref="GI:13183003"
                     /translation="MTPQPSGAPTVQVTRETERSFPRASEDEVTCPTSAPPSPTRTRG
                     NCAEAEEGGCRGAPRKLRARRGGRSRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSLYALEPPAPHCGELG
                     SPGGSPGDWGSLYSPVSQAGSLSPAASLEERPGLLGATSSACLSPGSLAFSDFL"
BASE COUNT     1215 a   1500 c   1514 g   1111 t
ORIGIN

  Query Match           35.4%;  Score 517.4;  DB 9;  Length 5340;
  Best Local Similarity  67.4%;  Pred. No. 1.3e-105;
  Matches  826;  Conservative    0;  Mismatches  351;  Indels   48;  Gaps    5;

Qy         20 GTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCAGCCCGGCAGGCACGCT 79
Db       2535 GGCCCCAGGGCCCCGGCGCTGATTGGCCGGTGGCGCGGGCAGCAGCCGGGCAGGCACGCT 2594

Qy         80 CCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACGATTAGCAGCTCAGAAGT 139
Db       2595 CCTGGCCCGGGCGAAGCAGATAAAGCGTGCCAAGGGGCACACGACTTGCTGCTCAGGAAA 2654

Qy        140 CCCTCTGGGTCTCACCACTGC------------------ACAGAGGCCGAGGACCCCCTC 181
Db       2655 TCCCTGCGGTCTCACCGCCGCGCCTCGAGAGAGAGCGTGACAGAGGCCTCGGACCCCATT 2714

Qy        182 CGAGCTTCTTTGCTGCCTCCAGACGCAATTTACT-CCAGGCGAGGGCGCCTGCAGCTCAG 240
Db       2715 CTCTCTTCTTTTTCTCCTTTGGGGCTGGGGCAACTCCCAGGCGGGGGCGCCTGCAGCTCAG 2774

Qy        241 CAAAACTTCGAAGCGAGCAG--------------AGGGGTTCAGCTATCCACCGCTGCT 285
Db       2775 CTGAACTTGGCGACCAGAAGCCCGCTGAGCTCCCCACGGCCCTCGCTGCTCATCGCTCTC 2834

Qy        286 TGACTCTGACCACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTA 345
Db       2835 TATTCTTTTGCGCCGGTAGAAAGGTAATATTTGGAGGCCTCCGAGGGACGGGCAGGGGAA 2894

Qy        346 GGAAC-CTCCAAAGGGTAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGT 404
Db       2895 AGAGGGATCCTCTGACCCAGCGGGGGCTGGGAGGATGGCTGTTTTTGTTTTTTCCCACCT 2954

Qy        405 GACCTCTAAGTCAGAGACTG-------------TCACACCCCCCTTCCATTTTTTCCCAA 451
Db       2955 AGCCTCGGAATCGCGGACTGCGCCGTGACGGACTCAAACTTACCCTTCCCTCTGACCCCG 3014

Qy        452 CCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCA 511
Db       3015 CCGTAGGATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGA 3074

Qy        512 GCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAG 571
Db       3075 GCGGTCCTTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCCAG 3134

Qy        572 CCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATC 631
Db       3135 CCCCACTCGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCCCC 3194

Qy        632 GAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAG 691
Db       3195 GAGGAAGCTCCGGGCACGGCGCGGGGGACGCAGCCGGCCTAAGAGCGAGTTGGCACTGAG 3254

Qy        692 CAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAA 751
Db       3255 CAAGCAGCGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACAA 3314

Qy        752 CCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAA 811
Db       3315 CCTCAACTCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCGAA 3374

Qy        812 ACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGAC 871
Db       3375 GCTCACCAAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAAC 3434

Qy        872 GCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCT 931
Db       3435 GCTGCGCATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGA 3494

Qy        932 GGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCA 991
Db       3495 GCTGGGCAGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCA 3554

Qy        992 AGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAG 1051
Db       3555 GGCTGGCAGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCAC 3614

Qy       1052 CTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTGTGAAGGGCCCA 1111
Db       3615 CTCTTCCGCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTGTGAAAGGACCT 3674

Qy       1112 AACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAGCTGTCTGAAATGG 1171
Db       3675 GTCTGTCGCTGGGCTGTGGGTGCTAAGGGTAAGGGAGAGGGAGGGAGCCGGGAGCCGTAG 3734

Qy       1172 AAGGTAGTGGAGGCACTCGAGCATC 1196
Db       3735 AGGGTGGCCGACGGCGGCGGCCCTC 3759


RESULT 12
AL450311/c
LOCUS       AL450311              165110 bp    DNA     linear   PRI 12-JUL-2001
DEFINITION  Human DNA sequence from clone RP11-343J3 on chromosome 10, complete
            sequence.
ACCESSION   AL450311
VERSION     AL450311.11  GI:14626972
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165110)
  AUTHORS   Howden,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2001 this sequence version replaced gi:14575291.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
```

chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr10
RP11-343J3 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-343J3 The true left end of clone RP11-242G20 is at 139955 in this sequence. The true right end of clone RP11-404C6 is at 6588 in this sequence.

FEATURES Location/Qualifiers
source 1. .165110
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-343J3"
/clone_lib="RPCI-11.2"
repeat_region 7. .147
/note="THE1C repeat: matches 2. .142 of consensus"
repeat_region 151. .293
/note="MIR repeat: matches 2. .148 of consensus"
repeat_region 1136. .1187
/note="26 copies 2 mer gt 98% conserved"
repeat_region 2960. .3272
/note="AluSg repeat: matches 1. .310 of consensus"
repeat_region 4144. .4270
/note="AluSx repeat: matches 1. .134 of consensus"
repeat_region 4319. .4494
/note="AluSx repeat: matches 118. .293 of consensus"
repeat_region 5662. .6438
/note="L1MEc repeat: matches 272. .1095 of consensus"
repeat_region 6916. .8080
/note="L1MEc repeat: matches 1168. .2367 of consensus"
repeat_region 9934. .10087
/note="MIR repeat: matches 25. .184 of consensus"
repeat_region 10353. .10481
/note="MIR repeat: matches 65. .194 of consensus"
repeat_region 11077. .11372
/note="AluSx repeat: matches 1. .294 of consensus"
repeat_region 11600. .11639
/note="20 copies 2 mer tc 95% conserved"
repeat_region 12479. .12608
/note="65 copies 2 mer at 90% conserved"
repeat_region 13433. .13876
/note="MLT1C repeat: matches 1. .466 of consensus"
repeat_region 14372. .14576
/note="AluSq repeat: matches 1. .301 of consensus"
repeat_region 14577. .14633
/note="AluSq repeat: matches 243. .299 of consensus"
repeat_region 14843. .14943
/note="MER86 repeat: matches 9. .109 of consensus"
repeat_region 15404. .15613
/note="MIR repeat: matches 3. .213 of consensus"
repeat_region 16626. .16705
/note="MIR repeat: matches 60. .140 of consensus"
repeat_region 16928. .16993
/note="33 copies 2 mer gg 66% conserved"
repeat_region 17476. .17569
/note="MER81 repeat: matches 2. .114 of consensus"
repeat_region 17719. .18069
/note="L1MC4 repeat: matches 7617. .7977 of consensus"
repeat_region 18117. .18345
/note="MIR repeat: matches 7. .262 of consensus"
repeat_region 19169. .19400
/note="MIR repeat: matches 5. .239 of consensus"
repeat_region 19436. .19627
/note="L1MC5 repeat: matches 7728. .7917 of consensus"
repeat_region 19642. .19716
/note="MIR repeat: matches 48. .131 of consensus"
repeat_region 19969. .20010
/note="21 copies 2 mer tg 100% conserved"
repeat_region 20623. .20704
/note="41 copies 2 mer gt 85% conserved"
repeat_region 20738. .20773
/note="U2 repeat: matches 1. .36 of consensus"
repeat_region 20802. .20863
/note="MLT1J repeat: matches 1. .62 of consensus"
repeat_region 20889. .21263
/note="MLT1F repeat: matches 188. .541 of consensus"
repeat_region 21463. .21618
/note="MIR repeat: matches 46. .192 of consensus"
repeat_region 22019. .22326
/note="AluJb repeat: matches 1. .306 of consensus"
repeat_region 22381. .22564
/note="L2 repeat: matches 2453. .2629 of consensus"
repeat_region 22896. .23174
/note="MLT1J repeat: matches 117. .413 of consensus"
repeat_region 23215. .23346
/note="MIR repeat: matches 48. .188 of consensus"
repeat_region 23388. .23532
/note="L2 repeat: matches 2097. .2230 of consensus"
repeat_region 23533. .23837
/note="AluJb repeat: matches 1. .303 of consensus"
repeat_region 23838. .24137
/note="L2 repeat: matches 1754. .2097 of consensus"
repeat_region 24291. .24581
/note="AluSx repeat: matches 1. .300 of consensus"
repeat_region 24653. .24850
/note="MIR repeat: matches 1. .200 of consensus"
misc_feature 26189. .26685
/note="CpG island"
/evidence=not_experimental
misc_feature 29453. .30918
/note="CpG island"
/evidence=not_experimental
repeat_region 31084. .31156
/note="MIR repeat: matches 65. .138 of consensus"
repeat_region 31618. .31876
/note="AluJo repeat: matches 29. .275 of consensus"
repeat_region 32767. .32830
/note="MIR repeat: matches 76. .139 of consensus"
repeat_region 33050. .33178
/note="43 copies 3 mer tcc 72% conserved"
misc_feature 35112. .36201
/note="CpG island"
/evidence=not_experimental
repeat_region 36551. .36604
/note="27 copies 2 mer ac 94% conserved"
repeat_region 37585. .38254
/note="L1PA15 repeat: matches 5480. .6157 of consensus"
repeat_region 38265. .38561
/note="AluSx repeat: matches 1. .300 of consensus"
repeat_region 39214. .39423
/note="MIR repeat: matches 3. .219 of consensus"
repeat_region 39489. .39801
/note="AluSx repeat: matches 1. .292 of consensus"
repeat_region 40050. .40189
/note="MIR repeat: matches 1. .144 of consensus"
repeat_region 41057. .41290
/note="MIR repeat: matches 8. .240 of consensus"
repeat_region 41296. .41460
/note="L2 repeat: matches 2569. .2730 of consensus"
repeat_region 41506. .41944
/note="L2 repeat: matches 1916. .2416 of consensus"
repeat_region 42388. .42698
/note="L2 repeat: matches 1448. .1779 of consensus"
repeat_region 44193. .44579
/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region 44600. .44956
/note="LTR16A repeat: matches 90. .445 of consensus"
repeat_region 45240. .45300
/note="MER58A repeat: matches 37. .97 of consensus"
repeat_region 45798. .45909
/note="LTR41 repeat: matches 90. .192 of consensus"
repeat_region 46826. .46871

```
                     /note="23 copies 2 mer gt 97% conserved"
     repeat_region   50684. .50969
                     /note="AluSq repeat: matches 1. .287 of consensus"
     repeat_region   50980. .51291
                     /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   52222. .52519
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   54065. .54260
                     /note="L1M4 repeat: matches 3865. .4055 of consensus"
     repeat_region   54261. .54432
                     /note="FAM repeat: matches 2. .167 of consensus"
     repeat_region   54433. .54629
                     /note="L1M4 repeat: matches 3652. .3865 of consensus"
     repeat_region   54648. .54862
                     /note="LTR41 repeat: matches 11. .217 of consensus"
     repeat_region   54863. .55236
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   55237. .55700

  Query Match             35.3%;  Score 515.8;  DB 9;  Length 165110;
  Best Local Similarity   66.8%;  Pred. No. 2e-105;
  Matches  820;  Conservative    0;  Mismatches  357;  Indels   50;  Gaps    4;

Qy         20 GTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCAGCCCGGCAGGCACGCT 79
              | |||  || |||||  |||||||||||| ||||| | |||||||||| ||||||||||||
Db      30814 GGCCCCAGGGCCCCGGCGCTGATTGGCCGGTGGCGCGGGCAGCAGCCGGGCAGGCACGCT 30755

Qy         80 CCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACGATTAGCAGCTCAGAAGT 139
              |||||  ||||||  |||||||||||||||||| ||| ||||||| | || |||||| |
Db      30754 CCTGGCCCGGGCGAAGCAGATAAAGCGTGCCAAGGGGCACACGACTTGCTGCTCAGGAAA 30695

Qy        140 CCCTCTGGGTCTCACCACTGC------------------ACAGAGGCCGAGGACCCCCTC 181
               ||    |||||||||  | ||                  |||||||||  ||||||| |
Db      30694 TCCCTGCGGTCTCACCGCCGCGCCTCGAGAGAGAGCGTGACAGAGGCCTCGGACCCCATT 30635

Qy        182 CGAGCTTCTTTGCTGCCTCCAGACGCAATTTACT-CCAGGCGAGGGCGCCTGCAGCTCAG 240
              |   ||||||| || | |    | |       ||| ||||||| ||||||||||||||||
Db      30634 CTCTCTTCTTTTCTCCTTTGGGGCTGGGGCAACTCCCAGGCGGGGGCGCCTGCAGCTCAG 30575

Qy        241 CAAAACTTCGAAGCGAGCAG---------------AGGGGTTCAGCTATCCACCGCTGCT 285
              |  |||||  |   | || ||              | ||     |||   || ||||
Db      30574 CTGAACTTGGCGACCAGAAGCCCGCTGAGCTCCCCACGGCCCTCGCTGCTCATCGCTCTC 30515

Qy        286 TGACTCTGACCACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTA 345
              |      |  | || | ||     |  | ||| ||| ||   |      ||  |     |
Db      30514 TATTCTTTTGCGCCGGTAGAAAGGTAATATTTGGAGGCCTCCGAGGGACGGGCAGGGGAA 30455

Qy        346 GGAACCTCCAAAGGGTAGAAGAGGGGAGTGGGTGGGCGTACT----------------CT 389
              ||     |   |    |  ||||   ||  |   |  |                   ||
Db      30454 AGAGGGATCCTCTGACCCAGCGGGGGCTGGGAGGATGGCTGTTTTTGTTTTTTCCCACCT 30395

Qy        390 AGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTGTCACACCCCCCTTCCATTTTTTCCC 449
              || | ||        ||| |        ||      ||| ||   || | |  | |  |||
Db      30394 AGCCTCGGAATCGCGGACTGCGCCCAGTGACGGACTCAAACTTACCCTTCCCTCTGACCC 30335

Qy        450 AACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACC 509
                ||  |||||| ||||||| |||| || |||||||||  |||||||| |||  |||||
Db      30334 CGCCGTAGGATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACG 30275

Qy        510 CAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCT 569
              |||   |||| ||| ||||||||||  ||||||||  | |  |||  ||| |||| ||
Db      30274 GAGCGGTCCTTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCC 30215

Qy        570 AGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACA 629
              |||||||||| |  || | || ||||| | || || ||||  || | ||||||||| |
Db      30214 AGCCCCACTCGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCC 30155

Qy        630 TCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTG 689
              |||||||||||||  || |||||||| || |||| | |||| ||||||||||||||||||
Db      30154 CCGAGGAAGCTCCGGGCACGGCGCGGGGGACGCAGCCGGCCTAAGAGCGAGTTGGCACTG 30095

Qy        690 AGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCAC 749
              |||||||||||||| || ||||| ||||||||||||||||| |||||||| || ||||||

Db      30094 AGCAAGCAGCGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCAC 30035

Qy        750 AACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCC 809
              ||||| ||||| || ||||| || |||||||||||||||||||||||||| || |||||
Db      30034 AACCTCAACTCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCG 29975

Qy        810 AAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAG 869
              || || || |||||||||| |||||||||||||||||||||| ||||| ||||||||
Db      29974 AAGCTCACCAAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAA 29915

Qy        870 ACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCCTGTGGGGAG 929
              |||||||||||||||||||||||||| ||||  |  ||||| || | |||     | | |
Db      29914 ACGCTGCGCATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGG 29855

Qy        930 CTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCC 989
                |     ||     ||| || |||  ||| |||||||| ||  ||||||||||| |||
Db      29854 GAGCTGGGCAGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCC 29795

Qy        990 CAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCC 1049
              || ||||| |||||||| ||| | || ||  ||||||||    || || |||| || | ||
Db      29794 CAGGCTGGCAGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCC 29735

Qy       1050 AGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTGTGAAGGGCC 1109
              | ||   | ||||  ||  ||| ||||  ||||  |||||||| ||  |||||| || |
Db      29734 ACCTTTTCCGCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTGTGAAAGGAC 29675

Qy       1110 CAAACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAGCTGTCTGAAAT 1169
              |  | | | |||||| |||| |||    | |||||| | |   |||| |   |  |
Db      29674 CTGTCTGTCGCTGGGCTGTGGGTGCTAAGGGTAAGGGAGAGGGAGGGAGCCGGGAGCCGT 29615

Qy       1170 GGAAGGTAGTGGAGGCACTCGAGCATC 1196
              || ||| |  || |    ||  | ||
Db      29614 AGAGGGTGGCCGACGGCGGCGGCCCTC 29588


RESULT 13
AC021954/c
LOCUS       AC021954               173341 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-57E12 map 10, WORKING DRAFT
            SEQUENCE, 24 unordered pieces.
ACCESSION   AC021954
VERSION     AC021954.3  GI:7417809
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173341)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 10, clone RP11-57E12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173341)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 173341)
```

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Apr 5, 2000 this sequence version replaced gi:6984451.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L5931
Center clone name: 57_E_12
---------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161190 bases at least Q40
Consensus quality: 166837 bases at least Q30
Consensus quality: 168995 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 171041; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1574: contig of 1574 bp in length
* 1575 1674: gap of 100 bp
* 1675 3043: contig of 1369 bp in length
* 3044 3143: gap of 100 bp
* 3144 5889: contig of 2746 bp in length
* 5890 5989: gap of 100 bp
* 5990 8979: contig of 2990 bp in length
* 8980 9079: gap of 100 bp
* 9080 13674: contig of 4595 bp in length
* 13675 13774: gap of 100 bp
* 13775 18831: contig of 5057 bp in length
* 18832 18931: gap of 100 bp
* 18932 23526: contig of 4595 bp in length
* 23527 23626: gap of 100 bp
* 23627 27386: contig of 3760 bp in length
* 27387 27486: gap of 100 bp
* 27487 32572: contig of 5086 bp in length
* 32573 32672: gap of 100 bp
* 32673 38632: contig of 5960 bp in length
* 38633 38732: gap of 100 bp
* 38733 43735: contig of 5003 bp in length
* 43736 43835: gap of 100 bp
* 43836 49020: contig of 5185 bp in length
* 49021 49120: gap of 100 bp
* 49121 53660: contig of 4540 bp in length
* 53661 53760: gap of 100 bp
* 53761 59544: contig of 5784 bp in length
* 59545 59644: gap of 100 bp
* 59645 66395: contig of 6751 bp in length
* 66396 66495: gap of 100 bp
* 66496 74696: contig of 8201 bp in length
* 74697 74796: gap of 100 bp
* 74797 81326: contig of 6530 bp in length
* 81327 81426: gap of 100 bp
* 81427 89446: contig of 8020 bp in length
* 89447 89546: gap of 100 bp
* 89547 96459: contig of 6913 bp in length
* 96460 96559: gap of 100 bp
* 96560 106498: contig of 9939 bp in length
* 106499 106598: gap of 100 bp
* 106599 116986: contig of 10388 bp in length
* 116987 117086: gap of 100 bp
* 117087 128890: contig of 11804 bp in length
* 128891 128990: gap of 100 bp
* 128991 147290: contig of 18300 bp in length
* 147291 147390: gap of 100 bp
* 147391 173341: contig of 25951 bp in length.

FEATURES Location/Qualifiers

source 1. .173341
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/map="10"
/clone="RP11-57E12"
/clone_lib="RPCI-11 Human Male BAC"

misc_feature 1. .1574
/note="assembly_fragment"

misc_feature 1675. .3043
/note="assembly_fragment"

misc_feature 3144. .5889
/note="assembly_fragment"

misc_feature 5990. .8979
/note="assembly_fragment"

misc_feature 9080. .13674
/note="assembly_fragment"

misc_feature 13775. .18831
/note="assembly_fragment"

misc_feature 18932. .23526
/note="assembly_fragment"

misc_feature 23627. .27386
/note="assembly_fragment
clone_end:SP6
vector_side:left"

misc_feature 27487. .32572
/note="assembly_fragment"

misc_feature 32673. .38632
/note="assembly_fragment"

misc_feature 38733. .43735
/note="assembly_fragment"

misc_feature 43836. .49020
/note="assembly_fragment"

misc_feature 49121. .53660
/note="assembly_fragment"

misc_feature 53761. .59544
/note="assembly_fragment"

misc_feature 59645. .66395
/note="assembly_fragment"

misc_feature 66496. .74696
/note="assembly_fragment"

```
     misc_feature    74797. .81326
                     /note="assembly_fragment"
     misc_feature    81427. .89446
                     /note="assembly_fragment"
     misc_feature    89547. .96459
                     /note="assembly_fragment"
     misc_feature    96560. .106498
                     /note="assembly_fragment"
     misc_feature    106599. .116986
                     /note="assembly_fragment"
     misc_feature    117087. .128890
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    128991. .147290
                     /note="assembly_fragment"
     misc_feature    147391. .173341
                     /note="assembly_fragment"

  Query Match             35.1%;  Score 512.6;  DB 2;  Length 173341;
  Best Local Similarity   66.7%;  Pred. No. 1.1e-104;
  Matches  818;  Conservative    0;  Mismatches  359;  Indels   50;  Gaps    4;

Qy         20 GTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCAGCCCGGCAGGCACGCT 79
              | |||  || |||||  |||||||||||| ||||| | |||||||||| ||||||||||||
Db       7254 GGCCCCAGGGCCCCGGCGCTGATTGGCCGGTGGCGCGGGCAGCAGCCGGGCAGGCACGCT 7195

Qy         80 CCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACGATTAGCAGCTCAGAAGT 139
              ||||| ||||||  |||||||||||||||||| ||| ||||||| | || ||||||  |
Db       7194 CCTGGCCCGGGCGAAGCAGATAAAGCGTGCCAAGGGGCACACGACTTGCTGCTCAGGAAA 7135

Qy        140 CCCTCTGGGTCTCACCACTGC------------------ACAGAGGCCGAGGACCCCCTC 181
               ||    |||||||||  | ||                  |||||||||  |||||||  |
Db       7134 TCCCTGCGGTCTCACCGCCGCGCCTCGAGAGAGAGCGTGACAGAGGCCTCGGACCCCATT 7075

Qy        182 CGAGCTTCTTTGCTGCCTCCAGACGCAATTTACT-CCAGGCGAGGGCGCCTGCAGCTCAG 240
              |   ||||||| || | |   | |       ||| ||||||| ||||||||||||||||
Db       7074 CTCTCTTCTTTTCTCCTTTGGGGCTGGGGCAACTCCCAGGCGGGGGCGCCTGCAGCTCAG 7015

Qy        241 CAAAACTTCGAAGCGAGCAG---------------AGGGGTTCAGCTATCCACCGCTGCT 285
              |  |||||  |   | || ||               | ||     |||   || ||||
Db       7014 CTGAACTTGGCGACCAGAAGCCCGCTGAGCTCCCCACGGCCCTCGCTGCTCATCGCTCTC 6955

Qy        286 TGACTCTGACCACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTA 345
              |     |  | || | ||     |  | |||  || ||   |      ||  |     |
Db       6954 TATTCTTTTGCGCCGGTAGAAAGGTAATATTTGCAGGCCTCCGAGGGACGGGCAGGGGAA 6895

Qy        346 GGAACCTCCAAAGGGTAGAAGAGGGGAGTGGGTGGGCGTACT----------------CT 389
               ||     |   |   |  ||||  ||  |   |   |                   ||
Db       6894 AGAGGGATCCTCTGACCCAGCGGGGGCTGGGAGGATGGCTGTTTTTGTTTTTTCCCACCT 6835

Qy        390 AGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTGTCACACCCCCCTTCCATTTTTTCCC 449
              || | ||         ||| |         ||     ||| ||   || | |  | |||
Db       6834 AGCCTCGGAATCGCGGACTGCGCCCAGTGACGGACTCAAACTTACCCTTCCCTCTGACCC 6775

Qy        450 AACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACC 509
               ||  |||||| ||||||| |||| || |||||||||  |||||||| |||  |||||
Db       6774 CGCCGTAGGATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACG 6715

Qy        510 CAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCT 569
               |||   |||| ||| ||||||||||  ||||||||  | |  |||  ||| |||| ||
Db       6714 GAGCGGTCCTTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCC 6655

Qy        570 AGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACA 629
              ||||||||||  |  || | || ||||| | || || ||||  || | ||||||||||  |
Db       6654 AGCCCCACTCGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCC 6595

Qy        630 TCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTG 689
               |||||||||||||  || ||||||||  || ||||  | |||  ||||||||||||||||||
Db       6594 CCGAGGAAGCTCCGGGCACGGCGCGGGGGACGCAGCCGGCTTAAGAGCGAGTTGGCACTG 6535

Qy        690 AGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCAC 749
              |||||||||||||| || ||||| |||||||||||||||||| |||||||| || ||||||

Db       6534 AGCAAGCAGCGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCAC 6475

Qy        750 AACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCC 809
              ||||| ||||| || ||||| || |||||||||||||||||||||||||| || |||||
Db       6474 AACCTCAACTCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCG 6415

Qy        810 AAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAG 869
              || || || |||||||||| |||||||||||||||||||||| ||||| |||||||
Db       6414 AAGCTCACCAAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAA 6355

Qy        870 ACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCCTGTGGGGAG 929
              |||||||||||||||||||||||| ||||  |  ||||| || | |||     | | |
Db       6354 ACGCTGCGCATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGG 6295

Qy        930 CTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCC 989
              |    ||     ||| || |||  ||| |||||||| ||  ||||||||||||| |||
Db       6294 GAGCTGGGCAGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCC 6235

Qy        990 CAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCC 1049
              || ||||| |||||||| ||| | || ||  ||||||||   || || |||| || | ||
Db       6234 CAGGCTGGCAGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCC 6175

Qy       1050 AGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTGTGAAGGGCC 1109
              | ||   | ||||  ||  ||| ||||  ||||  |||||||| ||  |||||| || |
Db       6174 ACCTTTTCCGCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTGTGAAAGGAC 6115

Qy       1110 CAAACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAGCTGTCTGAAAT 1169
              |   | | ||||||  |||||  |||     | |||||||  | |    ||||  |    |
Db       6114 CTGTCTGTCGCTGGGCTGTGGGTGCTAAGGGTAAGGGAGAGGGAGGGAGCCGGGAGCCGT 6055

Qy       1170 GGAAGGTAGTGGAGGCACTCGAGCATC 1196
              || ||| |  || |    ||  | ||
Db       6054 AGAGGGTGGCCGACGGCGGCGGCCCTC 6028
```

```
RESULT 14
HSA133776
LOCUS       HSA133776              1330 bp    DNA     linear   PRI 19-JUN-1999
DEFINITION  Homo sapiens gene for neurogenin 3.
ACCESSION   AJ133776
VERSION     AJ133776.1  GI:5123782
KEYWORDS    bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ravassard,P., Icard-Liepkalns,C., Wiard,L., Julien,J.P. and
            Mallet,J.
  TITLE     The human neurogenin 3 homolog maps to chromosome 10q21.3 and its
            expression pattern is identical to that of its murine counterparts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1330)
  AUTHORS   Ravassard,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-1999) Ravassard P., Lgn, CNRS UMRC 9923, Hopital
            de la Pitie Salpetriere, Bat. CERVI, 83 Bd. de l'Hopital, 75013
            PARIS, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1330
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            1. .1330
                     /gene="ngn3"
     mRNA            join(<1. .157,321. .>1330)
                     /gene="ngn3"
     exon            1. .157
                     /gene="ngn3"
                     /number=1
     intron          158. .320
                     /gene="ngn3"
                     /number=1
```

```
     exon            321. .1330
                     /gene="ngn3"
                     /number=2
     CDS             322. .966
                     /gene="ngn3"
                     /function="early neurogenesis"
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="CAB45384.1"
                     /db_xref="GI:5123783"
                     /translation="MTPQPSGAPTVQVTRETERSFPRASEDEVTCPTSAPPSPTRTPG
                     NCAEAEEGGCRGAPRKLRARRGGRSRPKSELALSKQRRSRRKKANDRERNRMHDLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSLYALEPPAPHCGELG
                     SPGGPPGDWGSLYSPVSQAGSLSPAASLEERPGLLGATSSACLSPGSLAFSDFL"
BASE COUNT      230 a    459 c    413 g    228 t
ORIGIN

  Query Match             31.3%;  Score 456.8;  DB 9;  Length 1330;
  Best Local Similarity   74.5%;  Pred. No. 6.1e-92;
  Matches  575;  Conservative    0;  Mismatches  197;  Indels    0;  Gaps    0;

Qy        425 TCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCC 484
              ||| ||   || | |  | |  |||  ||  |||||| ||||||| |||| || ||||||
Db        288 TCAAACTTACCCTTCCCTCTGACCCCGCCGTAGGATGACGCCTCAACCCTCGGGTGCGCC 347

Qy        485 CACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGT 544
              |||  |||||||| |||  |||||  |||   |||| ||| ||||||||||  |||||||
Db        348 CACTGTCCAAGTGACCCGTGAGACGGAGCGGTCCTTCCCCAGAGCCTCGGAAGACGAAGT 407

Qy        545 GCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGC 604
              |  | |  |||  ||| |||| || |||||||||| |  |||| || ||||| | || ||
Db        408 GACCTGCCCCACGTCCGCCCCGCCCAGCCCCACTCGCACACCGGGGAACTGCGCAGAGGC 467

Qy        605 AGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAA 664
              ||||  || | |||||||||| |  |||||||||||| || |||||||| || ||||
Db        468 GGAAGAGGGAGGCTGCCGAGGGGCCCCGAGGAAGCTCCGGGCACGGCGCGGGGGACGCAG 527

Qy        665 CAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAA 724
              | |||| ||||||||||||||||||||||||||||||||| || ||||| ||||||||||
Db        528 CCGGCCTAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGGAGTCGGCGAAAGAAGGCCAA 587

Qy        725 CGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGT 784
              |||||| |||||||| || |||||| |||| ||||| || ||||| || ||||||||||||
Db        588 CGACCGCGAGCGCAATCGAATGCACGACCTCAACTCGGCACTGGACGCCCTGCGCGGTGT 647

Qy        785 CCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCA 844
              |||||||||||||| || ||||| || || || |||||||||||| ||||||||||||||
Db        648 CCTGCCCACCTTCCCAGACGACGCGAAGCTCACCAAGATCGAGACGCTGCGCTTCGCCCA 707

Qy        845 CAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCC 904
              ||||||||| ||||| |||||||| |||||||||||||||||||||||||| ||||  |
Db        708 CAACTACATCTGGGCGCTGACTCAAACGCTGCGCATAGCGGACCACAGCTTGTACGCGCT 767

Qy        905 CGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTG 964
              |||||  || | |||     | | |  |    ||     ||| ||  ||  ||| |||||
Db        768 GGAGCCGCCGGCGCCGCACTGCGGGGAGCTGGGCAGCCCAGGCGGTCCCCCCGGGGACTG 827

Qy        965 GGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGA 1024
              ||| ||  |||||||||||| ||||| ||||||| ||| | || ||  ||||
Db        828 GGGGTCCCTCTACTCCCCAGTCTCCCAGGCTGGCAGCCTGAGTCCCGCCGCGTCGCTGGA 887

Qy       1025 GGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGT 1084
              ||||   || || |||| || | ||| |||  |  ||||  ||  ||| ||||  ||||
Db        888 GGAGCGACCCGGGCTGCTGGGGGCCACCTCTTCCGCCTGCTTGAGCCCAGGCAGTCTGGC 947

Qy       1085 GTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAG 1144
              ||||||||  ||  |||||| || ||   | | | |||||| |||||  |||    | |||
Db        948 TTTCTCAGATTTTCTGTGAAAGGACCTGTCTGTCGCTGGGCTGTGGGTGCTAAGGGTAAG 1007

Qy       1145 GGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATC 1196
              |||| | |  |||| |   |  | || ||| |  || |    ||  | ||
Db       1008 GGAGAGGGAGGGAGCCGGGAGCCGTAGAGGGTGGCCGACGGCGGCGGCCCTC 1059
```

```
RESULT 15
AC011010/c
LOCUS       AC011010              170896 bp    DNA     linear   HTG 16-MAR-2000
DEFINITION  Homo sapiens clone RP11-6P16, WORKING DRAFT SEQUENCE, 21 unordered
            pieces.
ACCESSION   AC011010
VERSION     AC011010.4  GI:7107881
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170896)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-6P16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170896)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2000 this sequence version replaced gi:6479051.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2916
                Center clone name: 6_P_16
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 114103 bases at least Q40
                Consensus quality: 141555 bases at least Q30
                Consensus quality: 158230 bases at least Q20
                Insert size: 154000; agarose-fp
                Insert size: 168896; sum-of-contigs
                Quality coverage: 3.6 in Q20 bases; agarose-fp
                Quality coverage: 3.3 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1744: contig of 1744 bp in length
            *     1745     1844: gap of 100 bp
            *     1845     2954: contig of 1110 bp in length
            *     2955     3054: gap of 100 bp
            *     3055     4839: contig of 1785 bp in length
            *     4840     4939: gap of 100 bp
```

```
            *     4940     7461: contig of 2522 bp in length
            *     7462     7561: gap of 100 bp
            *     7562    11912: contig of 4351 bp in length
            *    11913    12012: gap of 100 bp
            *    12013    15127: contig of 3115 bp in length
            *    15128    15227: gap of 100 bp
            *    15228    19176: contig of 3949 bp in length
            *    19177    19276: gap of 100 bp
            *    19277    24838: contig of 5562 bp in length
            *    24839    24938: gap of 100 bp
            *    24939    30632: contig of 5694 bp in length
            *    30633    30732: gap of 100 bp
            *    30733    37821: contig of 7089 bp in length
            *    37822    37921: gap of 100 bp
            *    37922    44638: contig of 6717 bp in length
            *    44639    44738: gap of 100 bp
            *    44739    49424: contig of 4686 bp in length
            *    49425    49524: gap of 100 bp
            *    49525    56517: contig of 6993 bp in length
            *    56518    56617: gap of 100 bp
            *    56618    65413: contig of 8796 bp in length
            *    65414    65513: gap of 100 bp
            *    65514    75659: contig of 10146 bp in length
            *    75660    75759: gap of 100 bp
            *    75760    86433: contig of 10674 bp in length
            *    86434    86533: gap of 100 bp
            *    86534    98763: contig of 12230 bp in length
            *    98764    98863: gap of 100 bp
            *    98864   109919: contig of 11056 bp in length
            *   109920   110019: gap of 100 bp
            *   110020   125264: contig of 15245 bp in length
            *   125265   125364: gap of 100 bp
            *   125365   141043: contig of 15679 bp in length
            *   141044   141143: gap of 100 bp
            *   141144   170896: contig of 29753 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170896
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-6P16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1744
                     /note="assembly_fragment"
     misc_feature    1845. .2954
                     /note="assembly_fragment"
     misc_feature    3055. .4839
                     /note="assembly_fragment"
     misc_feature    4940. .7461
                     /note="assembly_fragment"
     misc_feature    7562. .11912
                     /note="assembly_fragment"
     misc_feature    12013. .15127
                     /note="assembly_fragment"
     misc_feature    15228. .19176
                     /note="assembly_fragment"
     misc_feature    19277. .24838
                     /note="assembly_fragment"
     misc_feature    24939. .30632
                     /note="assembly_fragment"
     misc_feature    30733. .37821
                     /note="assembly_fragment"
     misc_feature    37922. .44638
                     /note="assembly_fragment"
     misc_feature    44739. .49424
                     /note="assembly_fragment"
     misc_feature    49525. .56517
                     /note="assembly_fragment"
     misc_feature    56618. .65413
                     /note="assembly_fragment"
     misc_feature    65514. .75659
                     /note="assembly_fragment"
     misc_feature    75760. .86433
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    86534. .98763
                     /note="assembly_fragment"
     misc_feature    98864. .109919
                     /note="assembly_fragment"
     misc_feature    110020. .125264
                     /note="assembly_fragment"
     misc_feature    125365. .141043
                     /note="assembly_fragment"
     misc_feature    141144. .170896
                     /note="assembly_fragment"
BASE COUNT    41547 a  41714 c  41241 g  44377 t   2017 others
ORIGIN

  Query Match             14.4%;  Score 210.6;  DB 2;  Length 170896;
  Best Local Similarity   56.3%;  Pred. No. 5.8e-37;
  Matches  282;  Conservative    0;  Mismatches  219;  Indels    0;  Gaps    0;

Qy        696 CAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTT 755
              |
Db      30733 CNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 30674

Qy        756 AACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTT 815
                                                  ||| || ||||| || ||
Db      30673 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCCAGACGACGCGAAGCTC 30614

Qy        816 ACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTG 875
              || |||||||||| |||||||||||||||||||||| ||||| |||||||| ||||||
Db      30613 ACCAAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAACGCTG 30554

Qy        876 CGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGA 935
              |||||||||||||||||| ||||  |  ||||| || | |||     | | |  |
Db      30553 CGCATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGAGCTG 30494

Qy        936 AGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCT 995
              ||     ||| || |||  ||| |||||||| ||  ||||||||||||| ||||| |||
Db      30493 GGCAGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCAGGCT 30434

Qy        996 GGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCC 1055
              || |||||||| ||| | || ||  ||||||||   || || |||| || | ||| ||
Db      30433 GGCAGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCACCTTT 30374

Qy       1056 CCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACA 1115
               |  |||| ||  ||| |||| |||| |||||||| ||  |||||| || ||   |
Db      30373 TCCGCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTGTGAAAGGACCTGTCT 30314

Qy       1116 GGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGG 1175
              | | |||||| ||||| |||    | |||||| | |   |||| |   |   | || ||
Db      30313 GTCGCTGGGCTGTGGGTGCTAAGGGTAAGGGAGAGGGAGGGAGCCGGGAGCCGTAGAGGG 30254

Qy       1176 TAGTGGAGGCACTCGAGCATC 1196
              | |  || |    ||  | ||
Db      30253 TGGCCGACGGCGGCGGCCCTC 30233


Search completed: January 26, 2004, 21:09:57
Job time : 5454 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 26, 2004, 18:38:36 ; Search time 424 Seconds
(without alignments)
9295.235 Million cell updates/sec

Title: US-09-595-947E-1
Perfect score: 1460
Sequence: 1 gcaggtagcgagaggagcag.....agagtgacctaatccagtgt 1460

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1458.4	99.9	1491	19	AAV42512	cDNA encoding a no
	2	900.6	61.7	1861	21	AAC61090	Murine neurogenin
	3	899.6	61.6	1860	24	AAD46872	Murine neurogenin
	4	819.8	56.2	5567	22	AAF27254	Mouse atonal homol
	5	682.4	46.7	861	22	AAF27266	Mouse neurogenin 3
	6	682.4	46.7	861	25	ABV75970	Mouse transcriptio
	7	652	44.7	804	19	AAV27050	Mouse neurogenin 3
	8	652	44.7	804	21	AAZ51981	Murine neurogenin-
	9	517.4	35.4	5340	21	AAC61089	Human neurogenin 3
	10	517.4	35.4	5340	24	AAD46871	Human neurogenin 3
c	11	229.6	15.7	592	24	ABQ49522	Oligonucleotide fo
	12	229.6	15.7	592	24	ABQ49523	Oligonucleotide fo
	13	192.2	13.2	592	24	ABQ49524	Oligonucleotide fo
c	14	192.2	13.2	592	24	ABQ49525	Oligonucleotide fo
	15	147.6	10.1	714	24	AAD46889	Human neurogenin 1
	16	147.6	10.1	1268	18	AAT74891	Human neurogenic d
	17	147.6	10.1	1268	19	AAV42932	DNA encoding human
	18	147.6	10.1	1268	25	ABS56390	Human bHLH family
	19	147.6	10.1	1665	24	AAD46888	Human neurogenin 1
	20	145.2	9.9	1527	19	AAV27045	Rat neurogenin 1 g
	21	145.2	9.9	1527	21	AAZ51976	Rat neurogenin-1 (
	22	140.4	9.6	738	19	AAV27046	Mouse neurogenin 1
	23	140.4	9.6	738	21	AAZ51977	Murine neurogenin-
	24	140.4	9.6	1332	19	AAV42938	DNA encoding murin
	25	140.4	9.6	1332	25	ABS56396	Mouse bHLH family
	26	140.4	9.6	1333	18	AAT74894	Mouse neurogenic d
	27	139.8	9.6	1385	19	AAV27049	Mouse neurogenin 2
	28	139.8	9.6	1385	21	AAZ51980	Murine neurogenin-
	29	139.8	9.6	1385	22	AAF27269	Mouse neurogenin 2
	30	139.8	9.6	1412	22	AAF27255	Mouse atonal homol
	31	139.8	9.6	1412	22	AAF27273	Mouse atonal homol
	32	139.2	9.5	790	22	AAF27264	Chicken atonal hom
	33	139.2	9.5	6123	24	AAD46890	Human neurogenin 2
	34	138.2	9.5	1074	22	AAF27263	Chicken atonal hom
	35	115.6	7.9	1312	19	AAV27047	Xenopus neurogenin
	36	115.6	7.9	1312	21	AAZ51978	Xenopus X-ngnr-1a
	37	109.8	7.5	1277	19	AAV27048	Xenopus neurogenin
	38	109.8	7.5	1277	21	AAZ51979	Xenopus X-ngnr-1b
c	39	106.6	7.3	778	24	ABQ16590	Oligonucleotide fo
	40	106.6	7.3	778	24	ABQ16591	Oligonucleotide fo
	41	96	6.6	1535	18	AAT74890	Human neurogenic d
	42	96	6.6	1535	19	AAV42931	DNA encoding human
	43	96	6.6	1535	25	ABS56389	Human bHLH family
	44	96	6.6	1550	21	AAA62681	Human NeuroD2 gene
	45	93.2	6.4	1951	18	AAT74893	Mouse neurogenic d

ALIGNMENTS

RESULT 1
AAV42512
ID AAV42512 standard; cDNA; 1491 BP.
XX
AC AAV42512;
XX
DT 05-OCT-1998 (first entry)
XX
DE cDNA encoding a novel BHLH protein designated RELAX.
XX
KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;
KW control; gene expression; transcriptional activator; targeting;
KW protein expression; central nervous system; CNS; treatment;
KW nervous system disorder; ss.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT CDS 459..1103
FT /*tag= a
FT /product= RELAX
XX
PN WO9827206-A2.
XX
PD 25-JUN-1998.
XX
PF 19-DEC-1997; 97WO-FR02368.
XX
PR 19-DEC-1996; 96FR-0015651.
XX
PA (RHON) RHONE-POULENC RORER SA.

```
XX
PI   Mallet J,  Ravassard P,  Icard-Liepkalns C;
XX
DR   WPI; 1998-362775/31.
DR   P-PSDB; AAW62991.
XX
PT   Basic helix-loop-helix polypeptide and related nucleic acid - with
PT   transcriptional activity, for targeting expression of genes to
PT   central nervous system and treatment of nervous disease
XX
PS   Claim 6; Page 20; 28pp; French.
XX
CC   The present sequence encodes a basic helix-loop-helix (BHLH) type
CC   protein, designated RELAX (Rat Embryonic Longitudinal Axis) protein.
CC   The protein is used to control and  participate in gene expression,
CC   by acting as transcriptional activator, strictly dependent on the
CC   presence of an intact E box (CANNTG), particularly for targeting
CC   expression of proteins to the central nervous system (CNS). The
CC   nucleic acid sequence can be used to treat nervous system disorders,
CC   and antisense sequences can be used to control mRNA transcription.
XX
SQ   Sequence 1491 BP; 307 A; 487 C; 413 G; 284 T; 0 other;

  Query Match             99.9%;  Score 1458.4;  DB 19;  Length 1491;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1459;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60

Qy         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120

Qy        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180

Qy        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240

Qy        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300

Qy        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360

Qy        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420

Qy        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480

Qy        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540

Qy        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600

Qy        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660

Qy        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720

Qy        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780

Qy        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840

Qy        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900

Qy        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960

Qy        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020

Qy       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080

Qy       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140

Qy       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200

Qy       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260

Qy       1261 GCTGTGCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTGTGCGTGCACAAAGGACATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320

Qy       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380

Qy       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440

Qy       1441 AGAGTGACCTAATCCAGTGT 1460
              ||||||||||||||||||||
Db       1441 AGAGTGACCTAATCCAGTGT 1460


RESULT 2
AAC61090
ID   AAC61090 standard; DNA; 1861 BP.
XX
AC   AAC61090;
XX
DT   05-FEB-2001  (first entry)
XX
DE   Murine neurogenin 3 (Ngn3) genomic DNA sequence.
XX
KW   Neurogenin 3; Ngn3; cellular differentiation; diabetes mellitus;
KW   islet cell precursor identification; mouse; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1093..1737
FT                   /*tag=  a
```

```
FT                   /product= "Ngn3"
FT                   /note= "Neurogenin 3"
XX
PN   WO200059936-A1.
XX
PD   12-OCT-2000.
XX
PF   28-MAR-2000; 2000WO-US08436.
XX
PR   06-APR-1999;   99US-0128180.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   German MS,  Lin J;
XX
DR   WPI; 2000-664989/64.
DR   P-PSDB; AAY85618.
XX
PT   Novel human neurogenin 3 polypeptides and polynucleotides encoding
PT   them, useful for diagnosis, prevention and treatment of diabetes
PT   mellitus and to identify individuals at risk of diabetes -
XX
PS   Claim 18; Page 49-50; 54pp; English.
XX
CC   The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
CC   protein AAY85617. The Ngn3 gene is located at chromosome position
CC   10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
CC   protein sequences, and includes an antibody recognising the Ngn3 protein.
CC   Also included in the invention is a method for identifying an islet cell
CC   precursor, the method involves analysing a cell for the expression of the
CC   Ngn3 gene product, where detection of the product is indicative of an
CC   islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
CC   reagent for detecting (in a subject) a predisposition to a defect in
CC   pancreatic islet cell function or formation associated with a defect in
CC   Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
CC   precursor cells expressing Ngn3, and to alter cellular differentiation in
CC   culture in vivo to produce new beta-cells to treat patients with diabetes
CC   mellitus. The present sequence represents the murine Ngn3 genomic DNA
CC   sequence.
XX
SQ   Sequence 1861 BP; 397 A; 560 C; 537 G; 367 T; 0 other;

  Query Match             61.7%;  Score 900.6;  DB 21;  Length 1861;
  Best Local Similarity   86.2%;  Pred. No. 4.9e-235;
  Matches 1094;  Conservative    0;  Mismatches  129;  Indels   46;  Gaps    7;

Qy          4 GGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCA 63
              ||   || | || |||| ||||||||||||   ||||||||||||||||  |||||||||
Db        594 GGGCTCGCGTGGCGCAGCCCCTGGGCCCCCTCCGCTGATTGGCCCGTGGTGCAGGCAGCA 653

Qy         64 GCCCGGCAGGCACGCTCCTGGTC-CGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 122
              |||||||||||||||||||| |  ||||||||||||||||||||||||||||||||||||
Db        654 GCCCGGCAGGCACGCTCCTGGCCGGGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 713

Qy        123 ATT----AGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCA-CAGAGGCCGAGGACCC 177
              | |     |||||||||| |||||||||||||| ||||||| ||| || |||| |||
Db        714 ACTTGCATGCAGCTCAGAAATCCCTCTGGGTCTCATCACTGCAGCAGTGGTCGAGTACCT 773

Qy        178 CCTCCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCT 237
              |||| |||||| | | |  | ||||||||||||||||||||||||||||||||||| |
Db        774 CCTCGGAGCTTTTCTACGACTTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGTT 833

Qy        238 CAGCAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACC- 296
              |||| ||||||  || ||||||||| || |||||||||||| ||||||||| ||||||
Db        834 TAGCAGAACTTCAGAGGGAGCAGAGAGGCTCAGCTATCCACTGCTGCTTGACACTGACCC 893

Qy        297 --------------------------ACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTA 330
                                        ||| || ||||||| |||||||||| | |||| |
Db        894 TATCCACTGCTGCTTGTCACTGACTGACCTGCTGCTCTCTATTCTTTTGAGTCGGGAG-A 952

Qy        331 ACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGGAGTGGGT----------GG 380
              |||||||||| || |||| |||||||||| || |||||| | | |              ||
Db        953 ACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGGCGCGCGGGGTGTGTGTGGG 1012

Qy        381 GCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTG--TCACACCCCCCTTC 438
              |  |||||| ||||| |||| |||||||||||||||||||| |||   ||||| | |||||
Db       1013 GGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGGCTGGCACACACACACCTTC 1072

Qy        439 CATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGT 498
              |||||||||||||| ||||||||||||||||||||||||||||| ||||||||||||||
Db       1073 CATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGT 1132

Qy        499 CCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 558
              ||| |||||| || ||||| ||||||||||||||||||||||||||||||||||||||||
Db       1133 CCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 1192

Qy        559 CCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACT 618
              |||||||||||||||||||||| |||| |||||||||||||||||||||  |||||||
Db       1193 CCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACT 1252

Qy        619 GCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCG 678
              ||||||||| |||||||||||||| || ||||||||||||||||||||||||||||||||
Db       1253 GCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCG 1312

Qy        679 AGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCA 738
              ||||||||| ||||| |||||| ||||||||||||||||||||||| || ||||||||||
Db       1313 AGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCA 1372

Qy        739 ACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 798
              | |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||
Db       1373 ATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 1432

Qy        799 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGG 858
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       1433 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGG 1492

Qy        859 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGC 918
              |||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db       1493 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGC 1552

Qy        919 CCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACT 978
              ||||||| |||||||| ||||| ||||| ||||||| ||| ||||||||||||||||||||
Db       1553 CCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACT 1612

Qy        979 CCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCC 1038
              ||||||| |||||||| |||| |||||||||||| |||||||||||||| ||||||||||
Db       1613 CCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCC 1672

Qy       1039 TGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCT 1098
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1673 TGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCT 1732

Qy       1099 TGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAG 1158
              ||||||| | ||   | ||| ||||| |||||| ||| |  ||||||||||   |||||
Db       1733 TGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTGGAAAGGGAGGGGACCAGAG 1792

Qy       1159 CTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGCCCCTTCTGGCTTTCATTA 1218
              | ||||| | ||| |||||||||||||| ||| |||||||||| ||||||||||||| ||
Db       1793 CCGTCTGGAGTGGGAGGTAGTGGAGGCTCTCAAGCATCTCGCCTCTTCTGGCTTTCACTA 1852

Qy       1219 GTCAGGTCC 1227
                |  | |||
Db       1853 CTTGGATCC 1861


RESULT 3
AAD46872
ID   AAD46872 standard; DNA; 1860 BP.
XX
AC   AAD46872;
XX
DT   27-JAN-2003  (first entry)
XX
DE   Murine neurogenin 3 (Ngn3) gene.
XX
KW   Transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
```

KW type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW islet cell; cell therapy; neurogenin 3; Ngn3; murine; gene; ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 1093..1737
FT /*tag= a
FT /product= "Murine Ngn3 protein"
XX
PN WO200274045-A2.
XX
PD 26-SEP-2002.
XX
PF 20-MAR-2002; 2002WO-US11166.
XX
PR 20-MAR-2001; 2001US-0817360.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI German MS, Lin J;
XX
DR WPI; 2002-759853/82.
DR P-PSDB; AAE29278.
XX
PT Producing a mammalian islet cell for treating diabetes mellitus
PT comprises introducing into a mammalian cell a nucleic acid molecule
PT encoding neuroendocrine basic helix-loop-helix transcription factor -
XX
PS Example 3; Page 89-90; 108pp; English.
XX
CC The invention relates to a method for producing a mammalian islet cell.
CC The method comprising introducing into a mammalian cell a nucleic acid
CC molecule encoding an islet transcription factor for expression of the
CC islet transcription factor in the cell and for production of islet cell
CC phenotype in the cell. The islet transcription factor is a neuroendocrine
CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC for treating type 2 diabetes mellitus and for replacing beta cells lost
CC to autoimmune destruction in individuals with type 1 diabetes. The method
CC is useful in cell therapy. The present sequence is murine neurogenin 3
CC (Ngn3) gene.
XX
SQ Sequence 1860 BP; 397 A; 559 C; 537 G; 367 T; 0 other;

Query Match 61.6%; Score 899.6; DB 24; Length 1860;
Best Local Similarity 86.2%; Pred. No. 9.2e-235;
Matches 1093; Conservative 0; Mismatches 129; Indels 46; Gaps 7;

```
Qy          4 GGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCA 63
              ||   || | || |||| ||||||||||||   ||||||||||||||| |||||||||
Db        594 GGGCTCGCGTGGCGCAGCCCCTGGGCCCCCTCCGCTGATTGGCCCGTGGTGCAGGCAGCA 653

Qy         64 GCCCGGCAGGCACGCTCCTGGTC-CGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 122
              |||||||||||||||||||||| |  ||||||||||||||||||||||||||||||||||
Db        654 GCCCGGCAGGCACGCTCCTGGCCGGGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 713

Qy        123 ATT----AGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCA-CAGAGGCCGAGGACCC 177
              | |    |||||||||||| |||||||||||||||| ||||||| ||| || |||| |||
Db        714 ACTTGCATGCAGCTCAGAAATCCCTCTGGGTCTCATCACTGCAGCAGTGGTCGAGTACCT 773

Qy        178 CCTCCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCT 237
              |||| |||||| | | |  | |||||||||||||||||||||||||||||||||||| |
Db        774 CCTCGGAGCTTTTCTACGACTTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGTT 833

Qy        238 CAGCAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACC- 296
               |||| ||||||  || |||||||| || |||||||||| |||||||||| ||||||
Db        834 TAGCAGAACTTCAGAGGGAGCAGAGAGGCTCAGCTATCCACTGCTGCTTGACACTGACCC 893

Qy        297 -------------------------ACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTA 330
                                        ||| || ||||||| |||||||||| | |||| |
Db        894 TATCCACTGCTGCTTGTCACTGACTGACCTGCTGCTCTCTATTCTTTTGAGTCGGGAG-A 952

Qy        331 ACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGGAGTGGGT----------GG 380
              |||||||||| || |||| |||||||||| || |||||| | | |            ||
Db        953 ACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGGCGCGCGGGGTGTGTGTGGG 1012

Qy        381 GCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTG--TCACACCCCCCTTC 438
              | |||||| |||||  |||| |||||||||||||||||||| |||   ||||| | |||||
Db       1013 GGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGGCTGGCACACACACACCTTC 1072

Qy        439 CATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGT 498
              |||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||||||
Db       1073 CATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGT 1132

Qy        499 CCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 558
              ||| |||||| || ||||| ||||||||||||||||||||||||||||||||||||||||
Db       1133 CCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 1192

Qy        559 CCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACT 618
              |||||||||||||||||||||||| |||| ||||||||||||||||||||  |||||||
Db       1193 CCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACT 1252

Qy        619 GCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCG 678
              |||||||||| |||||||||||||| || || ||||||||||||||||||||||||||||
Db       1253 GCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCG 1312

Qy        679 AGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCA 738
              |||||||||| ||||| |||||| |||||||||||||||||||||| || ||||||||||
Db       1313 AGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCA 1372

Qy        739 ACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 798
              | |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||
Db       1373 ATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 1432

Qy        799 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGG 858
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       1433 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGG 1492

Qy        859 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGC 918
              |||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||||
Db       1493 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGC 1552

Qy        919 CCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACT 978
              ||||||| |||||||| ||||| ||||| ||||||| ||| ||||||||||||||||||||
Db       1553 CCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACT 1612

Qy        979 CCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCC 1038
              ||||||| |||||||| |||| |||||||||||| |||||||||||||| ||||||||||
Db       1613 CCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCC 1672

Qy       1039 TGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCT 1098
              |||||||||||||||||||||| |||||||||||  | ||||||||||||||||||||
Db       1673 TGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCT 1732

Qy       1099 TGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAG 1158
              ||||||| | ||    | ||| ||||| ||||| ||| |  ||||||||||    |||||
Db       1733 TGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTGGAAAGGGAGGGGACCAGAG 1792

Qy       1159 CTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGCCCCTTCTGGCTTTCATTA 1218
              | ||||| | ||| |||||||||||||| ||| |||||||||| |||||||||||||| ||
Db       1793 CCGTCTGGAGTGGGAGGTAGTGGAGGCTCTCAAGCATCTCGCCTCTTCTGGCTTTCACTA 1852

Qy       1219 GTCAGGTC 1226
              |  | ||
Db       1853 CTTGGATC 1860
```

RESULT 4
AAF27254
ID AAF27254 standard; cDNA; 5567 BP.
XX
AC AAF27254;
XX
DT 24-APR-2001 (first entry)
XX
DE Mouse atonal homologue 5 (ATOH5, Math4B) cDNA, SEQ ID NO:4.

```
XX
KW   Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW   hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW   cellular proliferation; cerebellar granule neuron; gene therapy;
KW   mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW   transgenic animal; ss.
XX
OS   Mus musculus.
XX
PN   WO200073764-A2.
XX
PD   07-DEC-2000.
XX
PF   01-JUN-2000; 2000WO-US15410.
XX
PR   01-JUN-1999;   99US-0137060.
PR   19-JAN-2000; 2000US-0176993.
XX
PA   (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI   Zoghbi HY,  Bellen H,  Birmingham N,  Hassan B,  Ben-Arie N;
XX
DR   WPI; 2001-032190/04.
DR   P-PSDB; AAB60350.
XX
PT   Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT   any of its homologs or orthologs, for the treatment of e.g. deafness,
PT   osteoarthritis and abnormal cell proliferation -
XX
PS   Disclosure; Page -; 142pp; English.
XX
CC   The invention relates to the use of atonal-associated nucleic acid or
CC   amino acid sequence, or any of its homologues or orthologues as
CC   therapeutic agents for the treatment of deafness, partial hearing loss,
CC   vestibular effects due to damage or loss of inner hair cells,
CC   osteoarthritis and abnormal cell proliferation. The invention also
CC   encompasses methods of screening for compounds which affect the
CC   expression of an atonal-associated nucleic acid sequence in an animal,
CC   and a transgenic animal in which an allele of a native atonal-associated
CC   gene is replaced by a heterologous nucleic acid sequence, thus
CC   inactivating the atonal-associated allele. The nucleic acids or proteins
CC   may be used in a method of treating an animal for hearing impairment,
CC   joint disease, balance disorders, abnormal cell proliferation, or other
CC   disease related to loss of a functional atonal-associated nucleic acid or
CC   protein. They may particularly be used to treat an animal with a
CC   deficiency in cerebellar granule neurons or their precursors, and may
CC   also be used in promoting mechanoreceptive cell growth and generating
CC   hair cells. The present sequence represents an atonal-associated nucleic
CC   acid sequence referred to in the invention.
CC   Note: The present sequence is not shown in the specification, but
CC   was obtained from GenBank.
XX
SQ   Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;

  Query Match           56.2%;  Score 819.8;  DB 22;  Length 5567;
  Best Local Similarity  87.1%;  Pred. No. 7.5e-213;
  Matches  997;  Conservative    0;  Mismatches  102;  Indels   46;  Gaps    7;

Qy          4 GGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCA 63
              ||   || | || |||| ||||||||||||   ||||||||||||||||  |||||||||
Db       4424 GGGCTCGCGTGGCGCAGCCCCTGGGCCCCCTCCGCTGATTGGCCCGTGGTGCAGGCAGCA 4483

Qy         64 GCCCGGCAGGCACGCTCCTGGTC-CGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 122
              |||||||||||||||||||| | ||||||||||||||||||||||||||||||||||||
Db       4484 GCCCGGCAGGCACGCTCCTGGCCGGGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 4543

Qy        123 ATT----AGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCA-CAGAGGCCGAGGACCC 177
               | |     |||||||||| |||||||||||||| ||||||| ||| || |||| |||
Db       4544 ACTTGCATGCAGCTCAGAAATCCCTCTGGGTCTCATCACTGCAGCAGTGGTCGAGTACCT 4603

Qy        178 CCTCCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCT 237
              |||| |||||| | | |  | ||||||||||||||||||||||||||||||||||| |
Db       4604 CCTCGGAGCTTTTCTACGACTTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGTT 4663

Qy        238 CAGCAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACC- 296
              |||| ||||||  || |||||||| || |||||||||||| |||||||||| ||||||
Db       4664 TAGCAGAACTTCAGAGGGAGCAGAGAGGCTCAGCTATCCACTGCTGCTTGACACTGACCC 4723

Qy        297 -------------------------ACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTA 330
                                        ||| || ||||||| |||||||||| | |||| |
Db       4724 TATCCACTGCTGCTTGTCACTGACTGACCTGCTGCTCTCTATTCTTTTGAGTCGGGAG-A 4782

Qy        331 ACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGGAGTG----------GGTGG 380
              |||||||||| || |||| |||||||||| || |||||| | |          |  ||
Db       4783 ACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGGCGCGCGGGGTGTGTGTGGG 4842

Qy        381 GCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTGTCACACCCC--CCTTC 438
              |  |||||| ||||| |||| |||||||||||||||||| ||| ||||| |   |||||
Db       4843 GGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGGCTGGCACACACACACCTTC 4902

Qy        439 CATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGT 498
              |||||||||||||| ||||||||||||||||||||||||||||| ||||||||||||||
Db       4903 CATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGT 4962

Qy        499 CCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 558
              ||| |||||| || ||||| ||||||||||||||||||||||||||||||||||||||||
Db       4963 CCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 5022

Qy        559 CCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACT 618
              |||||||||||||||||||||| |||| ||||||||||||||||||||  |||||||
Db       5023 CCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACT 5082

Qy        619 GCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCG 678
              ||||||||| |||||||||||||| || || ||||||||||||||||||||||||||||
Db       5083 GCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCG 5142

Qy        679 AGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCA 738
              ||||||||| ||||| |||||| |||||||||||||||||||||| || |||||||||
Db       5143 AGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCA 5202

Qy        739 ACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 798
              | |||||||||||||| ||||| ||||||||||||||||||||||||||||||||||||
Db       5203 ATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 5262

Qy        799 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGG 858
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       5263 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGG 5322

Qy        859 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGC 918
              |||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db       5323 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGC 5382

Qy        919 CCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACT 978
              ||||||| |||||||| |||| ||||| ||||||| ||| ||||||||||||||||||||
Db       5383 CCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACT 5442

Qy        979 CCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCC 1038
              ||||||| |||||||| |||| |||||||||||| |||||||||||||| ||||||||||
Db       5443 CCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCC 5502

Qy       1039 TGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCT 1098
              |||||||||||||||||||||||||||||| ||||||||||  | ||||||||||||||||||
Db       5503 TGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCT 5562

Qy       1099 TGTGA 1103
              |||||
Db       5563 TGTGA 5567


RESULT 5
AAF27266
ID   AAF27266 standard; cDNA; 861 BP.
XX
AC   AAF27266;
XX
DT   24-APR-2001  (first entry)
```

XX
DE Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.
XX
KW Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW cellular proliferation; cerebellar granule neuron; gene therapy;
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW transgenic animal; ss.
XX
OS Mus musculus.
XX
PN WO200073764-A2.
XX
PD 07-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15410.
XX
PR 01-JUN-1999; 99US-0137060.
PR 19-JAN-2000; 2000US-0176993.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
XX
DR WPI; 2001-032190/04.
DR P-PSDB; AAB60359.
XX
PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT any of its homologs or orthologs, for the treatment of e.g. deafness,
PT osteoarthritis and abnormal cell proliferation -
XX
PS Disclosure; Page -; 142pp; English.
XX
CC The invention relates to the use of atonal-associated nucleic acid or
CC amino acid sequence, or any of its homologues or orthologues as
CC therapeutic agents for the treatment of deafness, partial hearing loss,
CC vestibular effects due to damage or loss of inner hair cells,
CC osteoarthritis and abnormal cell proliferation. The invention also
CC encompasses methods of screening for compounds which affect the
CC expression of an atonal-associated nucleic acid sequence in an animal,
CC and a transgenic animal in which an allele of a native atonal-associated
CC gene is replaced by a heterologous nucleic acid sequence, thus
CC inactivating the atonal-associated allele. The nucleic acids or proteins
CC may be used in a method of treating an animal for hearing impairment,
CC joint disease, balance disorders, abnormal cell proliferation, or other
CC disease related to loss of a functional atonal-associated nucleic acid or
CC protein. They may particularly be used to treat an animal with a
CC deficiency in cerebellar granule neurons or their precursors, and may
CC also be used in promoting mechanoreceptive cell growth and generating
CC hair cells. The present sequence represents an atonal-associated nucleic
CC acid sequence referred to in the invention.
CC Note: The present sequence is not shown in the specification, but
CC was obtained from GenBank.
XX
SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

Query Match 46.7%; Score 682.4; DB 22; Length 861;
Best Local Similarity 89.7%; Pred. No. 1e-175;
Matches 771; Conservative 0; Mismatches 76; Indels 13; Gaps 3;

```
Qy        312 TTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGG 371
              |||||||||| | |||| |||||||||||| || |||| |||||||||| || ||||||
Db          2 TTCTTTTGAGTCGGGAG-AACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGG 60

Qy        372 AGTGGGT----------GGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGA 421
              | | |          |||  |||||| ||||| |||| ||||||||||||||||||||
Db         61 CGCGCGGGGTGTGTGTGGGGGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGG 120

Qy        422 CTG--TCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGAT 479
              |||   ||||| | |||||||||||||||||| ||||||||||||||||||||||||
Db        121 CTGGCACACACACACCTTCCATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGAT 180

Qy        480 GCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCAC 539
              |||| |||||||||||||||| |||||| || ||||| |||||||||||||||||||||
Db        181 GCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCAC 240

Qy        540 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCC 599
              |||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db        241 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCC 300

Qy        600 GAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGG 659
              |||||||||  |||||||||||||||| |||||||||||||| || || |||||||||
Db        301 GAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGG 360

Qy        660 CGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAG 719
              |||||||||||||||||||||||||||||||| ||||| |||||| ||||||||||||||||
Db        361 CGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAG 420

Qy        720 GCCAACGACCCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGC 779
              ||||| || |||||||||||| |||||||||||||| ||||| ||||||||||||||||||
Db        421 GCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGC 480

Qy        780 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 839
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 540

Qy        840 GCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAC 899
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAT 600

Qy        900 GGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGC 959
              ||||| |||||||||||||||||||| |||||||| |||| ||||| ||||||| |||
Db        601 GGCCCGGAGCCCCCTGTGCCCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGG 660

Qy        960 GACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCA 1019
              |||||||||||||||||||||||||| |||||||| |||| |||||||||||| ||||||
Db        661 GACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCA 720

Qy       1020 TTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACC 1079
              |||||||| |||||||||||||||||||||||||||||||| ||||||||||| |
Db        721 TTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCA 780

Qy       1080 CTGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCA 1139
              |||||||||||||||||||||||||| | ||   | ||| ||||| |||||| ||| |
Db        781 CTGGTGTTCTCAGACTTCTTGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTG 840

Qy       1140 GAAAGGGAGGGAGTCAGAGC 1159
              ||||||||||   || |||
Db        841 GAAAGGGAGGGGACCACAGC 860
```

RESULT 6
ABV75970
ID ABV75970 standard; cDNA; 861 BP.
XX
AC ABV75970;
XX
DT 11-FEB-2003 (first entry)
XX
DE Mouse transcription factor neurogenin 3 cDNA.
XX
KW Mouse; transcription factor; neurogenin 3; ngn3; stem cell;
KW differentiation; beta-cell; insulin; diabetes; hyperglycaemia;
KW glucose intolerance; antidiabetic; hypoglycaemic; gene therapy;
KW gene; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 160..804
FT /*tag= a
FT /product= "Mouse ngn3"
XX
PN WO200286107-A2.
XX
PD 31-OCT-2002.

```
XX
PF   19-APR-2002; 2002WO-EP04362.
XX
PR   19-APR-2001; 2001US-284531P.
XX
PA   (DEVE-) DEVELOGEN ENTWICKLUNGSBIOLOGISCHE FORSCH.
PA   (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.
XX
PI   Wobus AM,  St-Onge L,  Blyszczuk P,  Hoffmann U;
XX
DR   WPI; 2003-075629/07.
XX
PT   Differentiating stem cells into insulin-producing cells useful for
PT   treating pancreatic diseases, by culturing stem cells in suitable
PT   medium and activating gene involved in beta-cell differentiation -
XX
PS   Disclosure; Page 58-59; 62pp; English.
XX
CC   The present sequence is that of cDNA encoding the murine
CC   basic helix-loop-helix transcription factor neurogenin 3 (ngn3),
CC   a gene which is required for the specification of the early
CC   endocrine precursor in the pancreatic epithelium and which is
CC   down-regulated once endocrine differentiation begins.  The invention
CC   provides a claimed method for differentiating stem cells (especially
CC   embryonic, adult or somatic stem cells and primordial germ cells)
CC   into insulin-producing cells.  This involves culturing stem cells in
CC   a suitable medium and activating at least one gene involved in
CC   beta-cell differentiation.  Preferred genes including Pdx1, Pax4,
CC   Pax6 and ngn3 (see ABV75967-70).  Gene activation comprises the
CC   delivery of the gene into stem cells using a viral delivery
CC   system, or the delivery of a protein product of the gene into stem
CC   cells.  The insulin-producing cells can be transplanted into
CC   animals or human for treatment of pancreatic diseases, metabolic
CC   syndrome and metabolic disorders with impaired glucose levels such
CC   as diabetes, hyperglycaemia and impaired glucose tolerance
CC   (claimed).  The cells can also be used to identify compounds which
CC   stimulate beta-cell differentiation, insulin secretion or glucose
CC   responsiveness.  Differentiated beta-cells can be used to study the
CC   toxic and other effects of exogenous compounds on beta-cell
CC   function.  In an example from the invention, Pax6 cDNA was inserted
CC   into expression vector pACCMV.pLpA under the control of the
CC   cytomegalovirus promoter.
XX
SQ   Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

  Query Match            46.7%;  Score 682.4;  DB 25;  Length 861;
  Best Local Similarity  89.7%;  Pred. No. 1e-175;
  Matches  771;  Conservative    0;  Mismatches   76;  Indels   13;  Gaps    3;

Qy        312 TTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGG 371
              |||||||||| | |||| |||||||||||| || |||| |||||||||| || ||||||
Db          2 TTCTTTTGAGTCGGGAG-AACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGG 60

Qy        372 AGTGGGT----------GGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGA 421
              | | |          ||| |||||| ||||| |||| ||||||||||||||||||||
Db         61 CGCGCGGGGTGTGTGTGGGGGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGG 120

Qy        422 CTG--TCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGAT 479
              |||   ||||| | ||||||||||||||||||| ||||||||||||||||||||||||
Db        121 CTGGCACACACACACCTTCCATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGAT 180

Qy        480 GCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCAC 539
              |||| |||||||||||||||| |||||| || ||||| ||||||||||||||||||||||
Db        181 GCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCAC 240

Qy        540 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCC 599
              |||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db        241 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCC 300

Qy        600 GAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGG 659
              ||||||||||  |||||||||||||||||| |||||||||||||| || || ||||||||||
Db        301 GAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGG 360

Qy        660 CGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAG 719
              |||||||||||||||||||||||||||||| ||||| |||||| ||||||||||||||||
Db        361 CGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAG 420

Qy        720 GCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGC 779
              ||||| || |||||||||||| |||||||||||||| ||||| ||||||||||||||||||
Db        421 GCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGC 480

Qy        780 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 540

Qy        840 GCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAC 899
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAT 600

Qy        900 GGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGC 959
              ||||| |||||||||||||||||| |||||||| ||||| ||||| ||||||| |||
Db        601 GGCCCGGAGCCCCCTGTGCCCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGG 660

Qy        960 GACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCA 1019
              |||||||||||||||||||||||||||| |||||||| |||| |||||||||| ||||||
Db        661 GACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCA 720

Qy       1020 TTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACC 1079
              |||||||| |||||||||||||||||||||||||||||||||||||| |||||||||||  |
Db        721 TTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCA 780

Qy       1080 CTGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCA 1139
              |||||||||||||||||||||||||| | ||   | ||| ||||| |||||| ||| |
Db        781 CTGGTGTTCTCAGACTTCTTGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTG 840

Qy       1140 GAAAGGGAGGGAGTCAGAGC 1159
              ||||||||||   || |||
Db        841 GAAAGGGAGGGGACCACAGC 860


RESULT 7
AAV27050
ID   AAV27050 standard; cDNA; 804 BP.
XX
AC   AAV27050;
XX
DT   17-SEP-1998  (first entry)
XX
DE   Mouse neurogenin 3 gene.
XX
KW   ds; Mouse; neurogenin; expression vector; recombinant protein;
KW   antibody; neurogenesis.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             160..804
FT                   /*tag=  a
FT                   /product= "Mouse neurogenin 3"
XX
PN   WO9813491-A2.
XX
PD   02-APR-1998.
XX
PF   24-SEP-1997;   97WO-US17048.
XX
PR   17-SEP-1997;   97US-0932411.
PR   27-SEP-1996;   96US-0722570.
PR   12-NOV-1996;   96US-0030864.
PR   19-DEC-1996;   96US-0772009.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Anderson DJ,  Ma Q,  Sommer L;
XX
DR   WPI; 1998-230702/20.
```

```
DR   P-PSDB; AAW54947.
XX
PT   Mouse neurogenins, useful in neurogenesis - and recombinant nucleic
PT   acids and proteins derived from rat and Xenopus
XX
PS   Disclosure; Fig 9; 106pp; English.
XX
CC   The Mouse neurogenin 3 is one of several neurogenin proteins discussed
CC   in the present invention.  The neurogenin nucleic acids can be expressed
CC   in a host cell, transformed using an expression vector, to produce
CC   recombinant proteins.  The proteins and the antibodies raised against
CC   the proteins are useful in the study of neurogenesis.
XX
SQ   Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

  Query Match             44.7%;  Score 652;  DB 19;  Length 804;
  Best Local Similarity   90.9%;  Pred. No. 1.9e-167;
  Matches  731;  Conservative    0;  Mismatches   60;  Indels   13;  Gaps    3;

Qy        312 TTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGG 371
              |||||||||| | |||| |||||||||||| || |||| |||||||||| || ||||||
Db          2 TTCTTTTGAGTCGGGAG-AACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGG 60

Qy        372 AGTGGGT----------GGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGA 421
              | | |          ||| |||||| ||||| |||| ||||||||||||||||||
Db         61 CGCGCGGGGTGTGTGTGGGGGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGG 120

Qy        422 CTG--TCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGAT 479
              |||   ||||| | |||||||||||||||||| ||||||||||||||||||||||||
Db        121 CTGGCACACACACACCTTCCATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGAT 180

Qy        480 GCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCAC 539
              |||| |||||||||||||||| |||||| || ||||| ||||||||||||||||||||||
Db        181 GCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCAC 240

Qy        540 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCC 599
              |||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db        241 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCC 300

Qy        600 GAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGG 659
              |||||||||  |||||||||||||||||| |||||||||||||| || || |||||||||
Db        301 GAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGG 360

Qy        660 CGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAG 719
              |||||||||||||||||||||||||||||| ||||| |||||| ||||||||||||||||
Db        361 CGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAG 420

Qy        720 GCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGC 779
              ||||| || |||||||||| |||||||||||||| ||||| ||||||||||||||||||
Db        421 GCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGC 480

Qy        780 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 839
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 540

Qy        840 GCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAC 899
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAT 600

Qy        900 GGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGC 959
              ||||| |||||||||||||||||||| |||||||| ||||| ||||| ||||||| |||
Db        601 GGCCCGGAGCCCCCTGTGCCCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGG 660

Qy        960 GACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCA 1019
              |||||||||||||||||||||||||||||| |||||||| |||| |||||||||||| ||||||
Db        661 GACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCA 720

Qy       1020 TTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACC 1079
              |||||||| |||||||||||||||||||||||||||||||| |||||||||||  |
Db        721 TTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCA 780

Qy       1080 CTGGTGTTCTCAGACTTCTTGTGA 1103
              |||||||||||||||||||||||||
Db        781 CTGGTGTTCTCAGACTTCTTGTGA 804


RESULT 8
AAZ51981
ID   AAZ51981 standard; DNA; 804 BP.
XX
AC   AAZ51981;
XX
DT   04-JUL-2000  (first entry)
XX
DE   Murine neurogenin-3 (NGN3) nucleic acid sequence.
XX
KW   Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;
KW   Phox2a protein; neuronal subtype-specific marker; growth factor;
KW   neural differentiation; transplantation; neuronal dysfunction;
KW   optical nerve damage; auditory nerve damage, neurodegenerative disorder;
KW   neuroprotective; nootropic; anticonvulsant; antiParkinsonian; vulnerary;
KW   cerebroprotective; immunesuppressant; antiinfectious; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             160..804
FT                   /*tag=  a
FT                   /product= "Murine neurogenin-3 protein"
XX
PN   WO200009676-A2.
XX
PD   24-FEB-2000.
XX
PF   13-AUG-1999;   99WO-US18525.
XX
PR   14-AUG-1998;   98US-0096630.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Anderson DJ,  Lo L;
XX
DR   WPI; 2000-256250/22.
DR   P-PSDB; AAY70570.
XX
PT   Inducing non-neuronal cells to differentiate into neurons and for
PT   non-neuronal cells to express a neuronal subtype-specific marker,
PT   comprising contacting the non-neuronal cells with a vector containing
PT   neurogenin nucleic acid -
XX
PS   Claim 1; Fig 1J; 76pp; English.
XX
CC   The patent discloses a method for inducing non-neuronal cells (NNC) to
CC   differentiate into neurons and for NNCs to express a neuronal subtype
CC   -specific marker. Transformed host cells are used as sources of neuronal
CC   and other growth factors; in culture for screening compounds that
CC   modulate neural differentiation or as sources of recombinantly produced
CC   neurogenins and Phox2a proteins for use in transplantation. The cells
CC   also have a variety of in vivo uses, e.g. for transplantation at sites of
CC   neuronal dysfunction e.g. patients with hearing or vision loss due to
CC   optical or auditory nerve damage, brain or spinal cord injuries, and
CC   neurodegenerative disorders e.g. Alzheimer's disease. The present
CC   sequence encodes murine neurogenin-3 (NGN-3), a transcription factor.
CC   NNCs differentiate into neurons through the recombinant expression of a
CC   transcription factor that induces a core program of neurogenesis. Forced
CC   expression of murine NGN3 can elicit expression of at least some neuronal
CC   phenotypic markers even in NNCs.
XX
SQ   Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

  Query Match             44.7%;  Score 652;  DB 21;  Length 804;
  Best Local Similarity   90.9%;  Pred. No. 1.9e-167;
  Matches  731;  Conservative    0;  Mismatches   60;  Indels   13;  Gaps    3;

Qy        312 TTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGG 371
              |||||||||| | |||| |||||||||||| || |||| |||||||||| || ||||||
```

```
Db          2 TTCTTTTGAGTCGGGAG-AACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGG 60

Qy        372 AGTGGGT----------GGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGA 421
              | | |           |||  |||||| ||||| |||| ||||||||||||||||||||
Db         61 CGCGCGGGGTGTGTGTGGGGGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGG 120

Qy        422 CTG--TCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGAT 479
              |||   ||||| | ||||||||||||||||||| ||||||||||||||||||||||||||
Db        121 CTGGCACACACACACCTTCCATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGAT 180

Qy        480 GCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCAC 539
              |||| |||||||||||||||| |||||| || ||||| ||||||||||||||||||||||
Db        181 GCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCAC 240

Qy        540 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCC 599
              |||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db        241 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCC 300

Qy        600 GAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGG 659
              |||||||||| |||||||||||||||| |||||||||||||| || || |||||||||
Db        301 GAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGG 360

Qy        660 CGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAG 719
              |||||||||||||||||||||||||||| ||||| |||||| ||||||||||||||||||
Db        361 CGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAG 420

Qy        720 GCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGC 779
              ||||| || |||||||||| |||||||||||||| ||||| ||||||||||||||||||
Db        421 GCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGC 480

Qy        780 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 839
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 540

Qy        840 GCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAC 899
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAT 600

Qy        900 GGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGC 959
              ||||| |||||||||||||||||| |||||||| |||| |||||| ||||||| |||
Db        601 GGCCCGGAGCCCCCTGTGCCCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGG 660

Qy        960 GACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCA 1019
              |||||||||||||||||||||||||| |||||||| |||| |||||||||||| ||||||
Db        661 GACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCA 720

Qy       1020 TTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACC 1079
              |||||||| |||||||||||||||||||||||||||||||||| ||||||||||||   |
Db        721 TTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCA 780

Qy       1080 CTGGTGTTCTCAGACTTCTTGTGA 1103
              ||||||||||||||||||||||||
Db        781 CTGGTGTTCTCAGACTTCTTGTGA 804
```

RESULT 9
AAC61089
ID AAC61089 standard; DNA; 5340 BP.
XX
AC AAC61089;
XX
DT 05-FEB-2001 (first entry)
XX
DE Human neurogenin 3 (Ngn3) genomic DNA sequence.
XX
KW Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;
KW islet cell precursor identification; diabetes mellitus; human; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 3022..3666
FT /*tag= a
FT /product= "Ngn3"
FT /note= "Neurogenin 3"
XX
PN WO200059936-A1.
XX
PD 12-OCT-2000.
XX
PF 28-MAR-2000; 2000WO-US08436.
XX
PR 06-APR-1999; 99US-0128180.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI German MS, Lin J;
XX
DR WPI; 2000-664989/64.
DR P-PSDB; AAY85617.
XX
PT Novel human neurogenin 3 polypeptides and polynucleotides encoding
PT them, useful for diagnosis, prevention and treatment of diabetes
PT mellitus and to identify individuals at risk of diabetes -
XX
PS Claim 6; Page 46-48; 54pp; English.
XX
CC The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
CC protein AAY85617. The Ngn3 gene is located at chromosome position
CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
CC protein sequences, and includes an antibody recognising the Ngn3 protein.
CC Also included in the invention is a method for identifying an islet cell
CC precursor, the method involves analysing a cell for the expression of the
CC Ngn3 gene product, where detection of the product is indicative of an
CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
CC reagent for detecting (in a subject) a predisposition to a defect in
CC pancreatic islet cell function or formation associated with a defect in
CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
CC precursor cells expressing Ngn3, and to alter cellular differentiation in
CC culture in vivo to produce new beta-cells to treat patients with diabetes
CC mellitus.
XX
SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

Query Match 35.4%; Score 517.4; DB 21; Length 5340;
Best Local Similarity 67.4%; Pred. No. 1.8e-130;
Matches 826; Conservative 0; Mismatches 351; Indels 48; Gaps 5;

```
Qy          20 GTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCAGCCCGGCAGGCACGCT 79
               | |||  || |||||  |||||||||||| ||||| | |||||||||| ||||||||||
Db        2535 GGCCCCAGGGCCCCGGCGCTGATTGGCCGGTGGCGCGGGCAGCAGCCGGGCAGGCACGCT 2594

Qy          80 CCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACGATTAGCAGCTCAGAAGT 139
               ||||| ||||||  |||||||||||||||||| ||| ||||||| | || |||||| |
Db        2595 CCTGGCCCGGGCGAAGCAGATAAAGCGTGCCAAGGGGCACACGACTTGCTGCTCAGGAAA 2654

Qy         140 CCCTCTGGGTCTCACCACTGC------------------ACAGAGGCCGAGGACCCCCTC 181
                ||    ||||||||| | ||                  |||||||||  ||||||| |
Db        2655 TCCCTGCGGTCTCACCGCCGCGCCTCGAGAGAGAGCGTGACAGAGGCCTCGGACCCCATT 2714

Qy         182 CGAGCTTCTTTGCTGCCTCCAGACGCAATTTACT-CCAGGCGAGGGCGCCTGCAGCTCAG 240
               |  ||||||| || | |   | |    ||| |||||||| ||||||||||||||||||
Db        2715 CTCTCTTCTTTTCTCCTTTGGGGCTGGGGCAACTCCCAGGCGGGGGCGCCTGCAGCTCAG 2774

Qy         241 CAAAACTTCGAAGCGAGCAG---------------AGGGGTTCAGCTATCCACCGCTGCT 285
               | ||||| |   | || ||                | ||     |||   || ||||
Db        2775 CTGAACTTGGCGACCAGAAGCCCGCTGAGCTCCCCACGGCCCTCGCTGCTCATCGCTCTC 2834

Qy         286 TGACTCTGACCACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTA 345
               |      |  | || | ||      |  | ||| ||| ||   |      || |     |
Db        2835 TATTCTTTTGCGCCGGTAGAAAGGTAATATTTGGAGGCCTCCGAGGGACGGGCAGGGGAA 2894

Qy         346 GGAAC-CTCCAAAGGGTAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGT 404
               ||    |||   |       | |||  | | |   |  |  | | ||           |
Db        2895 AGAGGGATCCTCTGACCCAGCGGGGGCTGGGAGGATGGCTGTTTTTGTTTTTTCCCACCT 2954
```

```
Qy        405 GACCTCTAAGTCAGAGACTG-------------TCACACCCCCCTTCCATTTTTTCCCAA 451
              ||||  | ||  |||||                ||| ||   || | |  | | |||
Db       2955 AGCCTCGGAATCGCGGACTGCGCCGTGACGGACTCAAACTTACCCTTCCCTCTGACCCCG 3014

Qy        452 CCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCA 511
              ||  |||||| ||||||| |||| || |||||||||  |||||||| ||| |||||  |
Db       3015 CCGTAGGATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGA 3074

Qy        512 GCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAG 571
              ||   |||| ||| ||||||||||  ||||||||  | |  |||  ||| |||| || ||
Db       3075 GCGGTCCTTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCCAG 3134

Qy        572 CCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATC 631
              |||||||| |  || | || ||||| | || || ||||  || | ||||||||||  |  |
Db       3135 CCCCACTCGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCCCC 3194

Qy        632 GAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAG 691
              ||||||||||||  || |||||||| || |||| | |||| ||||||||||||||||||||
Db       3195 GAGGAAGCTCCGGGCACGGCGCGGGGGACGCAGCCGGCCTAAGAGCGAGTTGGCACTGAG 3254

Qy        692 CAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAA 751
              ||||||||||||  || ||||| ||||||||||||||||  |||||||| || ||||||||
Db       3255 CAAGCAGCGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACAA 3314

Qy        752 CCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAA 811
              |||  ||||| || |||| || |||||||||||||||||||||||||| || ||||| ||
Db       3315 CCTCAACTCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCGAA 3374

Qy        812 ACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGAC 871
              ||  || ||||||||||| |||||||||||||||||||||||| |||||  ||||||| ||
Db       3375 GCTCACCAAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAAC 3434

Qy        872 GCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCT 931
              |||||||||||||||||||||||| ||||  |  ||||| || | |||     | | |
Db       3435 GCTGCGCATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGA 3494

Qy        932 GGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCA 991
              |    ||     ||| || |||  ||| ||||||||  ||  ||||||||||||| |||||
Db       3495 GCTGGGCAGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCA 3554

Qy        992 AGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAG 1051
              |||||  |||||||| ||| | || ||  ||||||||   || || |||| || | |||
Db       3555 GGCTGGCAGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCAC 3614

Qy       1052 CTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTGTGAAGGGCCCA 1111
              |||  |  |||| ||  ||| ||||  ||||  |||||||| ||  |||||| || ||
Db       3615 CTCTTCCGCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTGTGAAAGGACCT 3674

Qy       1112 AACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAGCTGTCTGAAATGG 1171
                | | | |||||| ||||| |||     | |||||| | |   ||||  |    |  | |
Db       3675 GTCTGTCGCTGGGCTGTGGGTGCTAAGGGTAAGGGAGAGGGAGGGAGCCGGGAGCCGTAG 3734

Qy       1172 AAGGTAGTGGAGGCACTCGAGCATC 1196
              | |||  |  || |    ||  | ||
Db       3735 AGGGTGGCCGACGGCGGCGGCCCTC 3759


RESULT 10
AAD46871
ID   AAD46871 standard; DNA; 5340 BP.
XX
AC   AAD46871;
XX
DT   27-JAN-2003  (first entry)
XX
DE   Human neurogenin 3 (Ngn3) gene.
XX
KW   Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW   type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW   islet cell; cell therapy; neurogenin 3; Ngn3; chromosome 10q22.1-22.2;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             3022..3666
FT                   /*tag=  a
FT                   /product= "Human Ngn3 protein"
XX
PN   WO200274045-A2.
XX
PD   26-SEP-2002.
XX
PF   20-MAR-2002; 2002WO-US11166.
XX
PR   20-MAR-2001; 2001US-0817360.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   German MS,  Lin J;
XX
DR   WPI; 2002-759853/82.
DR   P-PSDB; AAE29277.
XX
PT   Producing a mammalian islet cell for treating diabetes mellitus
PT   comprises introducing into a mammalian cell a nucleic acid molecule
PT   encoding neuroendocrine basic helix-loop-helix transcription factor  -
XX
PS   Example 2; Page 87-88; 108pp; English.
XX
CC   The invention relates to a method for producing a mammalian islet cell.
CC   The method comprising introducing into a mammalian cell a nucleic acid
CC   molecule encoding an islet transcription factor for expression of the
CC   islet transcription factor in the cell and for production of islet cell
CC   phenotype in the cell. The islet transcription factor is a neuroendocrine
CC   basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC   for treating type 2 diabetes mellitus and for replacing beta cells lost
CC   to autoimmune destruction in individuals with type 1 diabetes. The method
CC   is useful in cell therapy. The present sequence is human neurogenin 3
CC   (Ngn3) gene. Ngn3 gene is located on chromosome 10q22.1-22.2.
XX
SQ   Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

  Query Match           35.4%;  Score 517.4;  DB 24;  Length 5340;
  Best Local Similarity  67.4%;  Pred. No. 1.8e-130;
  Matches  826;  Conservative    0;  Mismatches  351;  Indels   48;  Gaps    5;

Qy         20 GTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCAGCCCGGCAGGCACGCT 79
              | |||  || |||||  |||||||||| ||||| | |||||||||| ||||||||||
Db       2535 GGCCCCAGGGCCCCGGCGCTGATTGGCCGGTGGCGCGGGCAGCAGCCGGGCAGGCACGCT 2594

Qy         80 CCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACGATTAGCAGCTCAGAAGT 139
              ||||| ||||||  |||||||||||||||||| ||| ||||||| | || ||||||  |
Db       2595 CCTGGCCCGGGCGAAGCAGATAAAGCGTGCCAAGGGGCACACGACTTGCTGCTCAGGAAA 2654

Qy        140 CCCTCTGGGTCTCACCACTGC-----------------ACAGAGGCCGAGGACCCCCTC 181
              ||     ||||||||| | ||                    |||||||||  |||||| |
Db       2655 TCCCTGCGGTCTCACCGCCGCGCCTCGAGAGAGAGCGTGACAGAGGCCTCGGACCCCATT 2714

Qy        182 CGAGCTTCTTTGCTGCCTCCAGACGCAATTTACT-CCAGGCGAGGGCGCCTGCAGCTCAG 240
              |   ||||||| || | |   | |        ||| ||||||| ||||||||||||||||
Db       2715 CTCTCTTCTTTTCTCCTTTGGGGCTGGGGCAACTCCCAGGCGGGGGCGCCTGCAGCTCAG 2774

Qy        241 CAAAACTTCGAAGCGAGCAG--------------AGGGGTTCAGCTATCCACCGCTGCT 285
              |  ||||| |   | || ||                | ||     |||   || ||||
Db       2775 CTGAACTTGGCGACCAGAAGCCCGCTGAGCTCCCCACGGCCCTCGCTGCTCATCGCTCTC 2834

Qy        286 TGACTCTGACCACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTA 345
              |    |   | || | ||     |  | ||| ||| ||   |       ||  |     |
Db       2835 TATTCTTTTGCGCCGGTAGAAAGGTAATATTTGGAGGCCTCCGAGGGACGGGCAGGGGAA 2894

Qy        346 GGAAC-CTCCAAAGGGTAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGT 404
              ||    |||   |      | |||  | | |   |  |  | | ||          |
Db       2895 AGAGGGATCCTCTGACCCAGCGGGGGCTGGGAGGATGGCTGTTTTTGTTTTTTCCCACCT 2954

Qy        405 GACCTCTAAGTCAGAGACTG-------------TCACACCCCCCTTCCATTTTTTCCCAA 451
```

```
Db       2955 AGCCTCGGAATCGCGGACTGCGCCGTGACGGACTCAAACTTACCCTTCCCTCTGACCCCG 3014

Qy        452 CCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCA 511

Db       3015 CCGTAGGATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGA 3074

Qy        512 GCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAG 571

Db       3075 GCGGTCCTTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCCAG 3134

Qy        572 CCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATC 631

Db       3135 CCCCACTCGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCCCC 3194

Qy        632 GAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAG 691

Db       3195 GAGGAAGCTCCGGGCACGGCGCGGGGGACGCAGCCGGCCTAAGAGCGAGTTGGCACTGAG 3254

Qy        692 CAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAA 751

Db       3255 CAAGCAGCGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACAA 3314

Qy        752 CCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAA 811

Db       3315 CCTCAACTCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCGAA 3374

Qy        812 ACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGAC 871

Db       3375 GCTCACCAAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAAC 3434

Qy        872 GCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCT 931

Db       3435 GCTGCGCATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGA 3494

Qy        932 GGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCA 991

Db       3495 GCTGGGCAGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCA 3554

Qy        992 AGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAG 1051

Db       3555 GGCTGGCAGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCAC 3614

Qy       1052 CTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTGTGAAGGGCCCA 1111

Db       3615 CTCTTCCGCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTGTGAAAGGACCT 3674

Qy       1112 AACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAGCTGTCTGAAATGG 1171

Db       3675 GTCTGTCGCTGGGCTGTGGGTGCTAAGGGTAAGGGAGAGGGAGGGAGCCGGGAGCCGTAG 3734

Qy       1172 AAGGTAGTGGAGGCACTCGAGCATC 1196

Db       3735 AGGGTGGCCGACGGCGGCGGCCCTC 3759


RESULT 11
ABQ49522/c
ID   ABQ49522 standard; DNA; 592 BP.
XX
AC   ABQ49522;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 36113.
XX
KW   Human; cytosine methylation; 5''-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 592 BP; 81 A; 59 C; 201 G; 251 T; 0 other;

  Query Match             15.7%;  Score 229.6;  DB 24;  Length 592;
  Best Local Similarity   67.6%;  Pred. No. 2e-52;
  Matches  322;  Conservative    0;  Mismatches  154;  Indels    0;  Gaps    0;

Qy        425 TCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCC 484

Db        477 TCAAACTTACCCTTCCCTCTAACCCCGCCGTAAAATAACGCCTCAACCCTCGAATACGCC 418

Qy        485 CACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGT 544

Db        417 CACTATCCAAATAACCCGTAAAACGAAACGATCCTTCCCCAAAACCTCGAAAAACGAAAT 358

Qy        545 GCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGC 604

Db        357 AACCTACCCCACGTCCGCCCCGCCCAACCCCACTCGCACACGAAAAAACTACGCAAAAAC 298

Qy        605 AGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAA 664

Db        297 GAAAAAAAAAAACTACCGAAAAACCCCGAAAAAACTCCGAACACGACGCGAAAAACGCAA 238

Qy        665 CAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAA 724

Db        237 CCGACTTAAAAACGAATTAACACTAAACAAACAACGACGAAATCGACGAAAAAAAACCAA 178

Qy        725 CGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGT 784

Db        177 CGACCGCGAACGCAATCGAATACACAACCTCAACTCGACACTAAACGCCCTACGCGATAT 118

Qy        785 CCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCA 844
```

```
                ||| ||||||||||||  | |||||  ||||| || || |||||  || || ||||||||||||
Db        117 CCTACCCACCTTCCCAAACGACGCGAAACTCACCAAAATCGAAACGCTACGCTTCGCCCA 58

Qy        845 CAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900
              |||||||||  |   | || |||||  ||||||  || ||||||  ||| ||||
Db         57 CAACTACATCTAAACGCTAACTCAAACGCTACGCATAACGAACCACAACTTATACG 2
```

RESULT 12
ABQ49523
ID ABQ49523 standard; DNA; 592 BP.
XX
AC ABQ49523;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36114.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;

  Query Match   15.7%; Score 229.6; DB 24; Length 592;
  Best Local Similarity 67.6%; Pred. No. 2e-52;
  Matches 322; Conservative   0; Mismatches 154; Indels   0; Gaps   0;

```
Qy        425 TCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCC 484
              |||  ||   ||  | |  | |  |||  ||  |  ||  ||||||| |||| | || ||||
Db        116 TCAAACTTACCCTTCCCTCTAACCCCGCCGTAAAATAACGCCTCAACCCTCGAATACGCC 175

Qy        485 CACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGT 544
              |||  |||||| |  |||   | ||  | | | |||| |||  | |||||  |  |||||  |
Db        176 CACTATCCAAATAACCCGTAAAACGAAACGATCCTTCCCCAAAACCTCGAAAAACGAAAT 235

Qy        545 GCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGC 604
                |   |||  ||| |||| || | ||||||||  |  ||  |   ||| | |  || |
Db        236 AACCTACCCCACGTCCGCCCCGCCCAACCCCACTCGCACACGAAAAAACTACGCAAAAAC 295

Qy        605 AGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAA 664
                ||  |   ||| |||| || |||  || ||||  | || |||| |   |||||
Db        296 GAAAAAAAAAAACTACCGAAAAACCCCGAAAAAACTCCGAACACGACGCGAAAAACGCAA 355

Qy        665 CAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAA 724
              | | |  || | ||| ||  |||| | ||| || |||||||  || || || ||  ||||
Db        356 CCGACTTAAAAACGAATTAACACTAAACAAACAACGACGAAATCGACGAAAAAAAACCAA 415

Qy        725 CGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGT 784
              |||||| || ||||| || || |||||||| |||||  | ||  | || || |||| | |
Db        416 CGACCGCGAACGCAATCGAATACACAACCTCAACTCGACACTAAACGCCCTACGCGATAT 475

Qy        785 CCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCA 844
              ||| ||||||||||||  | ||||| ||||| || || ||||| || || ||||||||||||
Db        476 CCTACCCACCTTCCCAAACGACGCGAAACTCACCAAAATCGAAACGCTACGCTTCGCCCA 535

Qy        845 CAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900
              |||||||||  |   | || |||||  ||||||  || ||||||  ||| ||||
Db        536 CAACTACATCTAAACGCTAACTCAAACGCTACGCATAACGAACCACAACTTATACG 591
```

RESULT 13
ABQ49524
ID ABQ49524 standard; DNA; 592 BP.
XX
AC ABQ49524;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36115.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 592 BP; 123 A; 59 C; 187 G; 223 T; 0 other;

Query Match 13.2%; Score 192.2; DB 24; Length 592;
Best Local Similarity 64.5%; Pred. No. 3.1e-42;
Matches 287; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```
Qy        456 AGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAA 515
              |||||| ||  | |    | || ||||   |   |  |||||   |  |||||  |||
Db        147 AGGATGACGTTTTAATTTTCGGGTGCGTTTATTGTTTAAGTGATTCGTGAGACGGAGCGG 206

Qy        516 CCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCC 575
                 |||    |||  |||||  |||||||  |  |||  |  | |    |   |||
Db        207 TTTTTTTTTAGAGTTTCGGAAGACGAAGTGATTTGTTTTACGTTCGTTTCGTTTAGTTTT 266

Qy        576 ACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGG 635
              | ||   ||| | || | |||   || || ||||  || |  || ||||||   ||||||
Db        267 ATTCGTATACGGGGGAATTGCGTAGAGGCGGAAGAGGGAGGTTGTCGAGGGGTTTCGAGG 326

Qy        636 AAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAG 695
              ||| | || |  |||||||| || || |   ||   |||||||||||||| | |||| |||
Db        327 AAGTTTCGGGTACGGCGCGGGGGACGTAGTCGGTTTAAGAGCGAGTTGGTATTGAGTAAG 386

Qy        696 CAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTT 755
               ||||||| || ||||| |||||||  ||||| || ||||| || || ||| | ||  ||
Db        387 TAGCGACGGAGTCGGCGAAAGAAGGTTAACGATCGCGAGCGTAATCGAATGTATAATTTT 446

Qy        756 AACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTT 815
              || || |   |||| |   |||||||||| ||   |  ||    || ||||| ||  ||
Db        447 AATTCGGTATTGGACGTTTTGCGCGGTGTTTTGTTTATTTTTTTAGACGACGCGAAGTTT 506

Qy        816 ACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTG 875
              |  |||||||||||  |||| ||||   | || || ||||||||  ||| | | ||| ||
Db        507 ATTAAGATCGAGACGTTGCGTTTCGTTTATAATTATATTTGGGCGTTGATTTAAACGTTG 566

Qy        876 CGCATAGCGGACCACAGCTTCTACG 900
              || ||||||||  | || || ||||
Db        567 CGTATAGCGGATTATAGTTTGTACG 591
```

RESULT 14
ABQ49525/c
ID ABQ49525 standard; DNA; 592 BP.
XX
AC ABQ49525;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36116.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 592 BP; 223 A; 187 C; 59 G; 123 T; 0 other;

Query Match 13.2%; Score 192.2; DB 24; Length 592;
Best Local Similarity 64.5%; Pred. No. 3.1e-42;
Matches 287; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```
Qy        456 AGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAA 515
              |||||| ||  | |    | || ||||   |   |  |||||   |  |||||  |||
Db        446 AGGATGACGTTTTAATTTTCGGGTGCGTTTATTGTTTAAGTGATTCGTGAGACGGAGCGG 387

Qy        516 CCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCC 575
                 |||    |||  |||||  |||||||  |  |||  |  | |    |   |||
Db        386 TTTTTTTTTAGAGTTTCGGAAGACGAAGTGATTTGTTTTACGTTCGTTTCGTTTAGTTTT 327

Qy        576 ACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGG 635
              | ||   ||| | || | |||   || || ||||  || |  || ||||||   ||||||
Db        326 ATTCGTATACGGGGGAATTGCGTAGAGGCGGAAGAGGGAGGTTGTCGAGGGGTTTCGAGG 267

Qy        636 AAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAG 695
              |||.| || |  |||||||| || || |   ||   |||||||||||||| | |||| |||
Db        266 AAGTTTCGGGTACGGCGCGGGGGACGTAGTCGGTTTAAGAGCGAGTTGGTATTGAGTAAG 207

Qy        696 CAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTT 755
               ||||||| || ||||| |||||||  ||||| || ||||| || || ||| | ||  ||
Db        206 TAGCGACGGAGTCGGCGAAAGAAGGTTAACGATCGCGAGCGTAATCGAATGTATAATTTT 147

Qy        756 AACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTT 815
              || || |   |||| |   |||||||||| ||   |  ||    || ||||| ||  ||
Db        146 AATTCGGTATTGGACGTTTTGCGCGGTGTTTTGTTTATTTTTTTAGACGACGCGAAGTTT 87
```

```
Qy        816 ACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTG 875
              |  ||||||||||||  |||| ||||   | || || ||||||||  ||| | | ||| ||
Db         86 ATTAAGATCGAGACGTTGCGTTTCGTTTATAATTATATTTGGGCGTTGATTTAAACGTTG 27

Qy        876 CGCATAGCGGACCACAGCTTCTACG 900
              || ||||||||  | || || ||||
Db         26 CGTATAGCGGATTATAGTTTGTACG 2
```

RESULT 15
AAD46889
ID AAD46889 standard; DNA; 714 BP.
XX
AC AAD46889;
XX
DT 27-JAN-2003 (first entry)
XX
DE Human neurogenin 1 (Ngn1) gene #2.
XX
KW Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW islet cell; cell therapy; neurogenin 1; Ngn1; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..714
FT /*tag= a
FT /product= "Human Ngn1 protein"
XX
PN WO200274045-A2.
XX
PD 26-SEP-2002.
XX
PF 20-MAR-2002; 2002WO-US11166.
XX
PR 20-MAR-2001; 2001US-0817360.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI German MS, Lin J;
XX
DR WPI; 2002-759853/82.
DR P-PSDB; AAE29280.
XX
PT Producing a mammalian islet cell for treating diabetes mellitus
PT comprises introducing into a mammalian cell a nucleic acid molecule
PT encoding neuroendocrine basic helix-loop-helix transcription factor -
XX
PS Disclosure; Page 94; 108pp; English.
XX
CC The invention relates to a method for producing a mammalian islet cell.
CC The method comprising introducing into a mammalian cell a nucleic acid
CC molecule encoding an islet transcription factor for expression of the
CC islet transcription factor in the cell and for production of islet cell
CC phenotype in the cell. The islet transcription factor is a neuroendocrine
CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC for treating type 2 diabetes mellitus and for replacing beta cells lost
CC to autoimmune destruction in individuals with type 1 diabetes. The method
CC is useful in cell therapy. The present sequence is human neurogenin 1
CC (Ngn1) gene.
XX
SQ Sequence 714 BP; 118 A; 287 C; 207 G; 102 T; 0 other;

Query Match 10.1%; Score 147.6; DB 24; Length 714;
Best Local Similarity 71.2%; Pred. No. 4.7e-30;
Matches 195; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```
Qy        647 GCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAG 706
              ||||||| | || || |   ||  |     |||| |   |||  |  || |||  | ||
Db        216 GCGGCGCCGCGGCCGGACGCGGGTCCGCTCCGAGGCGCTGCTGCACTCGCTGCGCAGGAG 275

Qy        707 CCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCT 766
              |||||||   |||||||||||| || ||||||||||||||||||||| | ||| | || ||
Db        276 CCGGCGCGTCAAGGCCAACGATCGCGAGCGCAACCGCATGCACAACTTGAACGCGGCCCT 335

Qy        767 GGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGA 826
              ||| || |||||| | || |||||| | ||||| || ||| |||| || || || |||||
Db        336 GGACGCACTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCTCACCAAAATCGA 395

Qy        827 GACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGA 886
              ||| |||||||||||| |||||||||| ||||| ||| |  |||| |||||| | |||||
Db        396 GACGCTGCGCTTCGCCTACAACTACATCTGGGCTCTGGCCGAGACACTGCGCCTGGCGGA 455

Qy        887 CCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC 920
              ||  |  |  |||  ||  ||| | || |
Db        456 TCAAGGGCTGCCCGGAGGCGGTGCCCGGGAGCGC 489
```

Search completed: January 26, 2004, 19:39:01
Job time : 430 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 26, 2004, 18:48:21 ; Search time 98 Seconds
(without alignments)
6575.708 Million cell updates/sec

Title: US-09-595-947E-1
Perfect score: 1460
Sequence: 1 gcaggtagcgagaggagcag.....agagtgacctaatccagtgt 1460

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	652	44.7	804	4	US-08-932-411A-19	Sequence 19, Appl
2	147.6	10.1	1268	1	US-08-910-973-12	Sequence 12, Appl
3	147.6	10.1	1268	4	US-09-499-227-12	Sequence 12, Appl
4	145.2	9.9	1527	4	US-08-722-570-12	Sequence 12, Appl
5	145.2	9.9	1527	4	US-08-932-411A-12	Sequence 12, Appl
6	140.4	9.6	738	4	US-08-722-570-13	Sequence 13, Appl
7	140.4	9.6	738	4	US-08-932-411A-13	Sequence 13, Appl
8	140.4	9.6	1333	1	US-08-910-973-21	Sequence 21, Appl
9	140.4	9.6	1333	4	US-09-499-227-21	Sequence 21, Appl
10	139.8	9.6	1385	4	US-08-932-411A-17	Sequence 17, Appl
11	121.8	8.3	310	1	US-08-552-142A-12	Sequence 12, Appl
12	115.6	7.9	1312	4	US-08-722-570-14	Sequence 14, Appl
13	115.6	7.9	1312	4	US-08-932-411A-15	Sequence 15, Appl
14	109.8	7.5	1277	4	US-08-722-570-15	Sequence 15, Appl
15	109.8	7.5	1277	4	US-08-932-411A-16	Sequence 16, Appl
16	96	6.6	1352	1	US-08-552-142A-10	Sequence 10, Appl
17	96	6.6	1535	1	US-08-910-973-10	Sequence 10, Appl
18	96	6.6	1535	4	US-09-499-227-10	Sequence 10, Appl
19	96	6.6	1550	3	US-09-234-332-3	Sequence 3, Appli
20	93.2	6.4	1462	1	US-08-552-142A-16	Sequence 16, Appl
21	93.2	6.4	1951	1	US-08-910-973-16	Sequence 16, Appl
22	93.2	6.4	1951	4	US-09-499-227-16	Sequence 16, Appl
23	92.4	6.3	2089	1	US-08-552-142A-1	Sequence 1, Appli
24	92.4	6.3	2089	1	US-08-910-973-1	Sequence 1, Appli
25	92.4	6.3	2089	4	US-09-499-227-1	Sequence 1, Appli
26	92.4	6.3	2089	5	PCT-US95-05741-1	Sequence 1, Appli
27	89	6.1	1560	1	US-08-552-142A-14	Sequence 14, Appl
28	89	6.1	1560	1	US-08-910-973-14	Sequence 14, Appl
29	89	6.1	1560	4	US-09-499-227-14	Sequence 14, Appl
30	89	6.1	2502	3	US-09-234-332-1	Sequence 1, Appli
31	85.8	5.9	1676	3	US-09-234-332-2	Sequence 2, Appli
32	83.4	5.7	524	1	US-08-552-142A-8	Sequence 8, Appli
33	83.4	5.7	524	1	US-08-910-973-8	Sequence 8, Appli
34	83.4	5.7	524	4	US-09-499-227-8	Sequence 8, Appli
35	83.4	5.7	524	5	PCT-US95-05741-8	Sequence 8, Appli
36	80	5.5	485	5	PCT-US95-05741-10	Sequence 10, Appl
37	73.6	5.0	1275	1	US-08-552-142A-3	Sequence 3, Appli
38	73.6	5.0	1275	1	US-08-910-973-3	Sequence 3, Appli
39	73.6	5.0	1275	4	US-09-499-227-3	Sequence 3, Appli
40	73.6	5.0	1275	5	PCT-US95-05741-3	Sequence 3, Appli
41	51.6	3.5	1635	3	US-09-234-332-4	Sequence 4, Appli
42	51.6	3.5	1635	4	US-09-702-705-1798	Sequence 1798, Ap
43	51.6	3.5	1635	4	US-09-736-457-1798	Sequence 1798, Ap
44	41	2.8	1785	5	PCT-US94-12912-1	Sequence 1, Appli
45	41	2.8	3636	1	US-07-753-520B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-932-411A-19
; Sequence 19, Application US/08932411A
; Patent No. 6566496
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qiufu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,411A
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/772,009
; FILING DATE: 19-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/722,570
; FILING DATE: 19-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 804 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 160..801
US-08-932-411A-19

```
  Query Match             44.7%;  Score 652;  DB 4;  Length 804;
  Best Local Similarity   90.9%;  Pred. No. 4.6e-163;
  Matches  731;  Conservative    0;  Mismatches   60;  Indels   13;  Gaps    3;

Qy        312 TTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGG 371
              |||||||||| | |||| |||||||||||| || |||| |||||||||||| || ||||||
Db          2 TTCTTTTGAGTCGGGAG-AACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGG 60

Qy        372 AGTGGGT----------GGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGA 421
              | | |           |||  |||||| |||| |||| ||||||||||||||||||||
Db         61 CGCGCGGGGTGTGTGTGGGGGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGG 120

Qy        422 CTG--TCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGAT 479
              |||   ||||| | ||||||||||||||||||||| ||||||||||||||||||||||||
Db        121 CTGGCACACACACACCTTCCATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGAT 180

Qy        480 GCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCAC 539
              |||| |||||||||||||||| |||||| || ||||| ||||||||||||||||||||||
Db        181 GCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCAC 240

Qy        540 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCC 599
              ||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db        241 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCC 300

Qy        600 GAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGG 659
              |||||||||  |||||||||||||||| |||||||||||||| || || |||||||||
Db        301 GAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGG 360

Qy        660 CGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAG 719
              |||||||||||||||||||||||||||| ||||| |||||| ||||||||||||||||
Db        361 CGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAG 420

Qy        720 GCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGC 779
              ||||| || |||||||||||| |||||||||||||| ||||| ||||||||||||||||
Db        421 GCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGC 480

Qy        780 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 540

Qy        840 GCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAC 899
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAT 600

Qy        900 GGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGC 959
              ||||| |||||||||||||||||| |||||||| ||||| ||||| |||||||| |||
Db        601 GGCCCGGAGCCCCCTGTGCCCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGG 660

Qy        960 GACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCA 1019
              |||||||||||||||||||||||| |||||||| |||| |||||||||||. ||||||
Db        661 GACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCA 720

Qy       1020 TTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACC 1079
              |||||||| ||||||||||||||||||||||||||||||||||| ||||||||||||  |
Db        721 TTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCA 780

Qy       1080 CTGGTGTTCTCAGACTTCTTGTGA 1103
              |||||||||||||||||||||||||
Db        781 CTGGTGTTCTCAGACTTCTTGTGA 804


RESULT 2
US-08-910-973-12
; Sequence 12, Application US/08910973
; Patent No. 5795723
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,973
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17532
;      FILING DATE:  30-October-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR-1-10958
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100; 206-224-0735 (direct)
;      TELEFAX:  206-225-0779
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1268 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  20A1 (neuroD3)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  55..768
US-08-910-973-12

  Query Match             10.1%;  Score 147.6;  DB 1;  Length 1268;
  Best Local Similarity   71.2%;  Pred. No. 1.1e-29;
  Matches  195;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        647 GCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAG 706
              |||||| | || || |   || |    |||| |  ||| | || ||| | ||
Db        270 GCGGCGCCGCGGCCGGACGCGGGTCCGCTCCGAGGCGCTGCTGCACTCGCTGCGCAGGAG 329

Qy        707 CCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCT 766
              |||||||   |||||||||||| || ||||||||||||||||||||| | ||| | || ||
Db        330 CCGGCGCGTCAAGGCCAACGATCGCGAGCGCAACCGCATGCACAACTTGAACGCGGCCCT 389

Qy        767 GGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGA 826
              ||| || |||||| | || |||||| | ||||| || ||| |||| || || || |||||
Db        390 GGACGCACTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCTCACCAAAATCGA 449

Qy        827 GACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGA 886
              ||| |||||||||||| |||||||||| ||||| ||| |  |||| |||||| | |||||
Db        450 GACGCTGCGCTTCGCCTACAACTACATCTGGGCTCTGGCCGAGACACTGCGCCTGGCGGA 509

Qy        887 CCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC 920
              ||  |  |   |||   ||   |||  | || |
Db        510 TCAAGGGCTGCCCGGAGGCGGTGCCCGGGAGCGC 543


RESULT 3
US-09-499-227-12
```

```
; Sequence 12, Application US/09499227
; Patent No. 6444463
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/499,227
;      FILING DATE:  05-August-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-May-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-May-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17532
;      FILING DATE:  30-October-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/910,973
;      FILING DATE:  07-August-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR-1-12742
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100; 206-224-0735 (direct)
;      TELEFAX:  206-225-0779
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1268 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  20A1 (neuroD3)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  55..768
US-09-499-227-12

  Query Match             10.1%;  Score 147.6;  DB 4;  Length 1268;
  Best Local Similarity   71.2%;  Pred. No. 1.1e-29;
  Matches  195;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        647 GCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAG 706
              ||||||| | || || |   ||  |      |||| |  |||  |  || |||  | ||
Db        270 GCGGCGCCGCGGCCGGACGCGGGTCCGCTCCGAGGCGCTGCTGCACTCGCTGCGCAGGAG 329

Qy        707 CCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCT 766
              |||||||   |||||||||||| || ||||||||||||||||||||| | ||| | || ||
Db        330 CCGGCGCGTCAAGGCCAACGATCGCGAGCGCAACCGCATGCACAACTTGAACGCGGCCCT 389

Qy        767 GGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGA 826
              ||| || |||||| | || |||||| | ||||| || ||| |||| || || || |||||
Db        390 GGACGCACTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCTCACCAAAATCGA 449

Qy        827 GACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGA 886
              ||| |||||||||||| |||||||||| ||||| ||| |  |||| ||||||  | |||||
Db        450 GACGCTGCGCTTCGCCTACAACTACATCTGGGCTCTGGCCGAGACACTGCGCCTGGCGGA 509

Qy        887 CCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC 920
              ||  |  |   |||   ||   |||  | || |
Db        510 TCAAGGGCTGCCCGGAGGCGGTGCCCGGGAGCGC 543


RESULT 4
US-08-722-570-12
; Sequence 12, Application US/08722570
; Patent No. 6555337
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  536S
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1527 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-12

  Query Match             9.9%;  Score 145.2;  DB 4;  Length 1527;
  Best Local Similarity   69.2%;  Pred. No. 4.8e-29;
  Matches  198;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        635 GAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAA 694
              | ||   ||   ||| ||  |||| |||    ||      |  ||||  |   |||  |
Db        456 GCAGGAGCGGCGGCGACGGCGAGGTCGCGCGCGGGTGCGGTCCGAGGCGCTGCTGCACTC 515

Qy        695 GCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCT 754
              || |||  | |||||  |||   |||||||||||  || ||||||||||| ||||| |||||
Db        516 GCTGCGGAGGAGCCGTCGCGTCAAGGCCAACGATCGCGAGCGCAACCGTATGCATAACCT 575

Qy        755 TAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACT 814
              |||  | |||||||| || |||||| | || ||||||  | |||||  || ||| |||| ||
Db        576 CAACGCTGCGCTGGACGCTCTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCT 635

Qy        815 TACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCT 874
              || ||||| ||||| |||||||||||| |||||||||| ||||| ||| || |||| ||
Db        636 CACCAAGATTGAGACGCTGCGCTTCGCCTACAACTACATCTGGGCCCTGGCTGAGACACT 695

Qy        875 GCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC 920
              |||| | || || ||  |  ||   ||   ||   |||  | || |
```

```
Db        696 GCGCCTGGCAGATCAAGGGCTCCCGGGGGGCGGTGCCCGGGAGCGC 741


RESULT 5
US-08-932-411A-12
; Sequence 12, Application US/08932411A
; Patent No. 6566496
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr Hohbach Test Albritton & Herbert LLP
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/932,411A
;      FILING DATE:  15-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/772,009
;      FILING DATE:  19-DEC-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,570
;      FILING DATE:  19-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902-3/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1527 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-932-411A-12

  Query Match             9.9%;  Score 145.2;  DB 4;  Length 1527;
  Best Local Similarity   69.2%;  Pred. No. 4.8e-29;
  Matches  198;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        635 GAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAA 694
              | ||   ||   ||| ||  |||| |||    ||     |  ||||  |   ||| |
Db        456 GCAGGAGCGGCGGCGACGGCGAGGTCGCGCGCGGGTGCGGTCCGAGGCGCTGCTGCACTC 515

Qy        695 GCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCT 754
              || |||  | ||||| |||   |||||||||| || ||||||||||| ||||| |||||
Db        516 GCTGCGGAGGAGCCGTCGCGTCAAGGCCAACGATCGCGAGCGCAACCGTATGCATAACCT 575

Qy        755 TAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACT 814
              ||| | |||||||| || |||||| | || |||||| | ||||| || ||| |||| ||
Db        576 CAACGCTGCGCTGGACGCTCTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCT 635

Qy        815 TACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCT 874
              || ||||| ||||| |||||||||||| |||||||||| ||||| ||| || |||| ||
Db        636 CACCAAGATTGAGACGCTGCGCTTCGCCTACAACTACATCTGGGCCCTGGCTGAGACACT 695

Qy        875 GCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC 920
              |||| | || || ||  |  ||   ||   ||   |||  | || |
Db        696 GCGCCTGGCAGATCAAGGGCTCCCGGGGGGCGGTGCCCGGGAGCGC 741


RESULT 6
US-08-722-570-13
; Sequence 13, Application US/08722570
; Patent No. 6555337
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  5365
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-13

  Query Match             9.6%;  Score 140.4;  DB 4;  Length 738;
  Best Local Similarity   68.2%;  Pred. No. 7.2e-28;
  Matches  195;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        635 GAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAA 694
              | ||   ||   | ||||  |||| |||    ||     |  ||||     ||| |
Db        207 GCAGGAACGGCGGAGGCGGCGAGGTCGCGCTCGGGTGCGGTCCGAGGCTCTGCTGCACTC 266

Qy        695 GCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCT 754
              | |||  | || || |||   || ||||||| || ||||||||||||||||||||||
Db        267 CCTGCGGAGGAGTCGTCGCGTCAAAGCCAACGATCGCGAGCGCAACCGCATGCACAACCT 326

Qy        755 TAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACT 814
              ||| | |||||||| || ||||| | || |||||| | ||||| || ||| |||| ||
Db        327 CAACGCTGCGCTGGACGCCTTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCT 386

Qy        815 TACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCT 874
              || ||||| ||||| |||||||||||| |||||||||| ||||| ||| || |||| ||
Db        387 CACCAAGATTGAGACGCTGCGCTTCGCCTACAACTACATCTGGGCCCTGGCTGAGACACT 446

Qy        875 GCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC 920
              |||| | || || ||  |  ||  |||   |    |||  | || |
Db        447 GCGCCTGGCAGATCAAGGGCTCCCCGGGGGCAGTGCCCGGGAGCGC 492


RESULT 7
US-08-932-411A-13
; Sequence 13, Application US/08932411A
```

```
; Patent No. 6566496
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr Hohbach Test Albritton & Herbert LLP
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/932,411A
;      FILING DATE:  15-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/772,009
;      FILING DATE:  19-DEC-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,570
;      FILING DATE:  19-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902-3/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..732
US-08-932-411A-13

  Query Match             9.6%;  Score 140.4;  DB 4;  Length 738;
  Best Local Similarity   68.2%;  Pred. No. 7.2e-28;
  Matches  195;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        635 GAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAA 694
              | ||   ||   | ||||  |||| |||    ||      |  ||||      |||  |
Db        207 GCAGGAACGGCGGAGGCGGCGAGGTCGCGCTCGGGTGCGGTCCGAGGCTCTGCTGCACTC 266

Qy        695 GCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCT 754
              | |||  | || || |||   || ||||||||| || ||||||||||||||||||||||
Db        267 CCTGCGGAGGAGTCGTCGCGTCAAAGCCAACGATCGCGAGCGCAACCGCATGCACAACCT 326

Qy        755 TAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACT 814
              ||| | |||||||| ||  ||||| | || |||||| | ||||| || ||| |||| ||
Db        327 CAACGCTGCGCTGGACGCCTTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCT 386

Qy        815 TACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCT 874
              || ||||| ||||| |||||||||||| |||||||||| |||| ||| || |||| ||
Db        387 CACCAAGATTGAGACGCTGCGCTTCGCCTACAACTACATCTGGGCCCTGGCTGAGACACT 446

Qy        875 GCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC 920
              |||| | || || ||  |  ||  |||   |    |||  | || |
Db        447 GCGCCTGGCAGATCAAGGGCTCCCCGGGGGCAGTGCCCGGGAGCGC 492
```

```
RESULT 8
US-08-910-973-21
; Sequence 21, Application US/08910973
; Patent No. 5795723
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,973
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17532
;      FILING DATE:  30-October-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR-1-10958
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100; 206-224-0735 (direct)
;      TELEFAX:  206-225-0779
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
;    IMMEDIATE SOURCE:
;      CLONE:  neuroD3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  101..835
US-08-910-973-21

  Query Match             9.6%;  Score 140.4;  DB 1;  Length 1333;
  Best Local Similarity   68.2%;  Pred. No. 8.6e-28;
  Matches  195;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        635 GAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAA 694
              | ||   ||   | ||||  |||| |||    ||      |  ||||      |||  |
Db        307 GCAGGAACGGCGGAGGCGGCGAGGTCGCGCTCGGGTGCGGTCCGAGGCTCTGCTGCACTC 366

Qy        695 GCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCT 754
              | |||  | || || |||   || ||||||||| || ||||||||||||||||||||||
Db        367 CCTGCGGAGGAGTCGTCGCGTCAAAGCCAACGATCGCGAGCGCAACCGCATGCACAACCT 426

Qy        755 TAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACT 814
              ||| | |||||||| ||  ||||| | || |||||| | ||||| || ||| |||| ||
Db        427 CAACGCTGCGCTGGACGCCTTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCT 486
```

```
Qy        815 TACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCT 874
              || ||||| ||||| |||||||||||||| |||||||||| ||||| ||| || |||| ||
Db        487 CACCAAGATTGAGACGCTGCGCTTCGCCTACAACTACATCTGGGCCCTGGCTGAGACACT 546

Qy        875 GCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC 920
              |||| | || || ||  |  ||  |||    |    |||  | || |
Db        547 GCGCCTGGCAGATCAAGGGCTCCCCGGGGGCAGTGCCCGGGAGCGC 592


RESULT 9
US-09-499-227-21
; Sequence 21, Application US/09499227
; Patent No. 6444463
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/499,227
;      FILING DATE:  05-August-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-May-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-May-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17532
;      FILING DATE:  30-October-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/910,973
;      FILING DATE:  07-August-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR-1-12742
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100; 206-224-0735 (direct)
;      TELEFAX:  206-225-0779
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
;    IMMEDIATE SOURCE:
;      CLONE:  neuroD3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  101..835
US-09-499-227-21

  Query Match             9.6%;  Score 140.4;  DB 4;  Length 1333;
  Best Local Similarity   68.2%;  Pred. No. 8.6e-28;
  Matches  195;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        635 GAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAA 694
              | ||   ||   | ||||  |||| |||    ||      |  ||||      |||  |
Db        307 GCAGGAACGGCGGAGGCGGCGAGGTCGCGCTCGGGTGCGGTCCGAGGCTCTGCTGCACTC 366

Qy        695 GCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCT 754
              | |||  | || || |||   || |||||||| || ||||||||||||||||||||||
Db        367 CCTGCGGAGGAGTCGTCGCGTCAAAGCCAACGATCGCGAGCGCAACCGCATGCACAACCT 426

Qy        755 TAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACT 814
              ||| | |||||||| ||  ||||| | || |||||| | ||||| || ||| |||| ||
Db        427 CAACGCTGCGCTGGACGCCTTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCT 486

Qy        815 TACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCT 874
              || ||||| ||||| |||||||||||| |||||||||| ||||| ||| || |||| ||
Db        487 CACCAAGATTGAGACGCTGCGCTTCGCCTACAACTACATCTGGGCCCTGGCTGAGACACT 546

Qy        875 GCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC 920
              |||| | || || ||  |  ||  |||    |    |||  | || |
Db        547 GCGCCTGGCAGATCAAGGGCTCCCCGGGGGCAGTGCCCGGGAGCGC 592


RESULT 10
US-08-932-411A-17
; Sequence 17, Application US/08932411A
; Patent No. 6566496
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr Hohbach Test Albritton & Herbert LLP
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/932,411A
;      FILING DATE:  15-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/772,009
;      FILING DATE:  19-DEC-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,570
;      FILING DATE:  19-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902-3/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1385 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  382..1170
US-08-932-411A-17

  Query Match             9.6%;  Score 139.8;  DB 4;  Length 1385;
  Best Local Similarity   78.1%;  Pred. No. 1.3e-27;
```

```
  Matches  168;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
              | |||  |    |||||||| |||| ||||||||||||||||||||| ||| |||||
Db        715 ACCCGCAGGCTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCG 774

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
              ||||| ||||||||||  || ||||||||||||||| || || ||||| || || ||||||
Db        775 CTGGACGCGCTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATC 834

Qy        825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCG 884
              ||||| ||||||||||||||||| ||||| |||| || ||  |||| |||||| | |||
Db        835 GAGACGCTGCGCTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCG 894

Qy        885 GACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCC 919
              |||||| ||  |  || |   | | ||  | | |
Db        895 GACCACTGCGCCGGCGCCGGTGGCCTCCAGGGGGC 929


RESULT 11
US-08-552-142A-12
; Sequence 12, Application US/08552142A
; Patent No. 5695995
;  GENERAL INFORMATION:
;    APPLICANT:  Weintraub, Harold M.
;    APPLICANT:  Lee, Jacqueline E.
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Hollenberg, Stanley M.
;    TITLE OF INVENTION:  Neurogenic Differentiation (NeuroD) Genes
;    TITLE OF INVENTION:  and Proteins
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/552,142A
;      FILING DATE:  02-NOV-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-MAY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Broderick, Thomas F.
;      REGISTRATION NUMBER:  31,332
;      REFERENCE/DOCKET NUMBER:  FHCR-1-8933
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100
;      TELEFAX:  206-225-0709
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  310 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  20A1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..310
US-08-552-142A-12

  Query Match             8.3%;  Score 121.8;  DB 1;  Length 310;
  Best Local Similarity   65.2%;  Pred. No. 4.6e-23;
  Matches  191;  Conservative    0;  Mismatches  101;  Indels    1;  Gaps    1;

Qy        619 GCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCG 678
              | | |||| ||  |    |    | | | ||||  |  |||   ||  |         ||
Db         16 GTCCAGGGGCACAGGACGACGAGCAGGAGAGGCGGCGGCGGCCGGACGCGNGTCCCTCCG 75

Qy        679 AGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCA 738
              ||  |   |||  | || |||  | || ||||||   |||||||||| || |||||||
Db         76 AGGCGCTGCTGCACNCGCTGCGCAGGAG-CGGCGCGTCAAGGCCAACGATCGCGAGCGCA 134

Qy        739 ACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 798
              |||||||||||||| | ||| | || ||||| || |||||| | || |||||| | ||||
Db        135 ACCGCATGCACAACTTGAACGCGGCCCTGGACGCACTGCGCAGCGTGCTGCCCTCGTTCC 194

Qy        799 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGG 858
              | || ||| |||| || || || ||||||| |||||| | |||| |||||||||| ||||
Db        195 CCGACGACACCAAGCTCACCAAAATCGAGAGCCTGCGTTNCGCCTACAACTACATCTGGG 254

Qy        859 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCC 911
              | ||| |  |||| ||||||     |   |  ||| |   |   | | ||||
Db        255 CTCTGGCCGAGACACTGCGCTGGCGGATNAAGGGCTGCCCGGAGGCGGTGCCC 307


RESULT 12
US-08-722-570-14
; Sequence 14, Application US/08722570
; Patent No. 6555337
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  5365
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1312 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-14

  Query Match             7.9%;  Score 115.6;  DB 4;  Length 1312;
  Best Local Similarity   67.4%;  Pred. No. 3.1e-21;
  Matches  163;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;
```

```
Qy        647 GCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAG 706
              | || || |||| |||      | ||||||||      |  ||   |||          |
Db        539 GAGGAGCCGAGGCCGCGCTCAGGGCAAGAGCGGAGAAACTGTGTTAAAGATCAAGAAGAC 598

Qy        707 CCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCT 766
              |||||||   || || ||| |||||||  | || ||||||||||||||| ||||| |||||
Db        599 CCGGCGCGTTAAAGCTAACAACCGGGAAAGGAATCGCATGCACAACCTGAACTCTGCGCT 658

Qy        767 GGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGA 826
              ||| | ||  | |  ||  ||||| | || || || || |||||||| || |||||  ||
Db        659 TGATTCCCTCAGGGAAGTGTTGCCCTCTTTACCTGAAGATGCCAAACTCACCAAGATAGA 718

Qy        827 GACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGA 886
              |||| ||||||| ||| |||||||||| ||||| || |   | ||  ||||| | |  ||
Db        719 GACCTTGCGCTTTGCCTACAACTACATCTGGGCTCTTAGCGAAACTTTGCGCCTTGGCGA 778

Qy        887 CC 888
              ||
Db        779 CC 780
```

RESULT 13
US-08-932-411A-15
; Sequence 15, Application US/08932411A
; Patent No. 6566496
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qiufu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,411A
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/772,009
; FILING DATE: 19-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/722,570
; FILING DATE: 19-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-932-411A-15

Query Match 7.9%; Score 115.6; DB 4; Length 1312;
Best Local Similarity 67.4%; Pred. No. 3.1e-21;
Matches 163; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```
Qy        647 GCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAG 706
              | || || |||| |||      | ||||||||      |  ||   |||          |
Db        539 GAGGAGCCGAGGCCGCGCTCAGGGCAAGAGCGGAGAAACTGTGTTAAAGATCAAGAAGAC 598

Qy        707 CCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCT 766
              |||||||   || || ||| |||||||  | || ||||||||||||||| ||||| |||||
Db        599 CCGGCGCGTTAAAGCTAACAACCGGGAAAGGAATCGCATGCACAACCTGAACTCTGCGCT 658

Qy        767 GGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGA 826
              ||| | ||  | |  ||  ||||| | || || || || |||||||| || |||||  ||
Db        659 TGATTCCCTCAGGGAAGTGTTGCCCTCTTTACCTGAAGATGCCAAACTCACCAAGATAGA 718

Qy        827 GACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGA 886
              |||| ||||||| ||| |||||||||| ||||| || |   | ||  ||||| | |  ||
Db        719 GACCTTGCGCTTTGCCTACAACTACATCTGGGCTCTTAGCGAAACTTTGCGCCTTGGCGA 778

Qy        887 CC 888
              ||
Db        779 CC 780
```

RESULT 14
US-08-722-570-15
; Sequence 15, Application US/08722570
; Patent No. 6555337
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qiufu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,570
; FILING DATE: 27-SEP-1996
; CLASSIFICATION: 5365
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
US-08-722-570-15

Query Match 7.5%; Score 109.8; DB 4; Length 1277;
Best Local Similarity 74.6%; Pred. No. 1e-19;
Matches 138; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
              | |||||||   || |||||  |||| ||| | || |||||||| |||| ||||  |||
Db        439 ACCCGGCGCGTTAAAGCCAATAACCGCGAGAGGAATCGCATGCACCACCTGAACTATGCG 498

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
```

```
                 || ||| | ||| | |  || || ||  | || || || ||||||||||| || |||||
Db        499 CTCGATTCTCTGAGGGAGGTTCTACCGTCATTACCCGAAGACGCCAAACTCACCAAGATA 558

Qy        825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCG 884
                 |||||| ||||||| |||||||||||||| ||||| || |   | ||  ||||| | ||
Db        559 GAGACCTTGCGCTTTGCCCACAACTACATCTGGGCTCTTAGCGAAACTTTGCGCCTGGCC 618

Qy        885 GACCA 889
                 |||||
Db        619 GACCA 623


RESULT 15
US-08-932-411A-16
; Sequence 16, Application US/08932411A
; Patent No. 6566496
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr Hohbach Test Albritton & Herbert LLP
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/932,411A
;      FILING DATE:  15-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/772,009
;      FILING DATE:  19-DEC-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,570
;      FILING DATE:  19-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902-3/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-932-411A-16

  Query Match             7.5%;  Score 109.8;  DB 4;  Length 1277;
  Best Local Similarity   74.6%;  Pred. No. 1e-19;
  Matches  138;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
                | |||||||   || |||||  |||| ||| | || |||||||| |||| ||||  |||
Db        439 ACCCGGCGCGTTAAAGCCAATAACCGCGAGAGGAATCGCATGCACCACCTGAACTATGCG 498

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
                 || ||| | ||| | |  || || ||  | || || || ||||||||||| || |||||
Db        499 CTCGATTCTCTGAGGGAGGTTCTACCGTCATTACCCGAAGACGCCAAACTCACCAAGATA 558

Qy        825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCG 884
                 |||||| ||||||| |||||||||||||| ||||| || |   | ||  ||||| | ||
Db        559 GAGACCTTGCGCTTTGCCCACAACTACATCTGGGCTCTTAGCGAAACTTTGCGCCTGGCC 618

Qy        885 GACCA 889
                 |||||
Db        619 GACCA 623


Search completed: January 26, 2004, 22:06:42
Job time : 100 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 26, 2004, 19:36:37 ; Search time 516 Seconds
(without alignments)
10119.471 Million cell updates/sec

Title: US-09-595-947E-1
Perfect score: 1460
Sequence: 1 gcaggtagcgagaggagcag.....agagtgacctaatccagtgt 1460

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2356869 seqs, 1788235258 residues

Total number of hits satisfying chosen parameters: 4713738

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	900.6	61.7	1861	9	US-09-817-360-3	Sequence 3, Appli
2	682.4	46.7	861	14	US-10-004-717-24	Sequence 24, Appl
3	582.6	39.9	645	14	US-10-004-717-4	Sequence 4, Appli
4	517.4	35.4	5340	9	US-09-817-360-1	Sequence 1, Appli
5	145.2	9.9	1527	8	US-08-722-570-12	Sequence 12, Appl
6	140.4	9.6	738	8	US-08-722-570-13	Sequence 13, Appl
7	139.8	9.6	1385	14	US-10-004-717-30	Sequence 30, Appl
8	139.8	9.6	1412	14	US-10-004-717-6	Sequence 6, Appli
9	139.8	9.6	1412	14	US-10-004-717-37	Sequence 37, Appl
10	139.2	9.5	790	14	US-10-004-717-20	Sequence 20, Appl
11	138.2	9.5	1074	14	US-10-004-717-18	Sequence 18, Appl
12	115.6	7.9	1312	8	US-08-722-570-14	Sequence 14, Appl
13	109.8	7.5	1277	8	US-08-722-570-15	Sequence 15, Appl
14	89	6.1	1099	13	US-10-413-358-27	Sequence 27, Appl
15	89	6.1	1211	13	US-10-413-358-26	Sequence 26, Appl
16	81.4	5.6	501	14	US-10-004-717-13	Sequence 13, Appl
17	78.4	5.4	485	14	US-10-004-717-59	Sequence 59, Appl
18	68.6	4.7	446	10	US-09-833-381-505	Sequence 505, App
19	65.8	4.5	938	14	US-10-004-717-39	Sequence 39, Appl
20	65.2	4.5	450	14	US-10-004-717-22	Sequence 22, Appl
21	64.4	4.4	651	10	US-09-749-728B-22	Sequence 22, Appl
22	64	4.4	2466	14	US-10-044-090-251	Sequence 251, App
23	63.4	4.3	65	13	US-09-908-975-4140	Sequence 4140, Ap
c 24	63.2	4.3	2106	13	US-10-027-632-97281	Sequence 97281, A
c 25	63.2	4.3	2106	13	US-10-027-632-97282	Sequence 97282, A
c 26	63.2	4.3	2106	13	US-10-027-632-97283	Sequence 97283, A
c 27	63.2	4.3	2106	14	US-10-027-632-97281	Sequence 97281, A
c 28	63.2	4.3	2106	14	US-10-027-632-97282	Sequence 97282, A
c 29	63.2	4.3	2106	14	US-10-027-632-97283	Sequence 97283, A
30	61.2	4.2	645	10	US-09-749-728B-24	Sequence 24, Appl
31	61.2	4.2	834	10	US-09-764-864-353	Sequence 353, App
32	61.2	4.2	1750	15	US-10-157-031-238	Sequence 238, App
33	60.8	4.2	1550	14	US-10-004-717-43	Sequence 43, Appl
34	60.8	4.2	1957	14	US-10-004-717-8	Sequence 8, Appli
35	59.8	4.1	65	13	US-09-908-975-30269	Sequence 30269, A
36	59.2	4.1	675	14	US-10-004-717-15	Sequence 15, Appl
37	58	4.0	907	14	US-10-004-717-65	Sequence 65, Appl
38	57.8	4.0	1065	14	US-10-004-717-1	Sequence 1, Appli
39	57.8	4.0	1572	14	US-10-004-717-57	Sequence 57, Appl
40	57.4	3.9	1056	14	US-10-004-717-10	Sequence 10, Appl
41	57.4	3.9	1393	14	US-10-004-717-45	Sequence 45, Appl
42	56.2	3.8	1034	13	US-10-029-386-22762	Sequence 22762, A
43	56.2	3.8	1217	13	US-10-027-632-101238	Sequence 101238,
44	56.2	3.8	1217	14	US-10-027-632-101238	Sequence 101238,
45	56.2	3.8	2993	15	US-10-125-237-56	Sequence 56, Appl

ALIGNMENTS

RESULT 1
US-09-817-360-3
; Sequence 3, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-817-360-3

Query Match 61.7%; Score 900.6; DB 9; Length 1861;
Best Local Similarity 86.2%; Pred. No. 4.5e-248;
Matches 1094; Conservative 0; Mismatches 129; Indels 46; Gaps 7;

```
Qy          4 GGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCA 63
              ||   || | || |||| |||||||||||   |||||||||||||||  |||||||||
Db        594 GGGCTCGCGTGGCGCAGCCCCTGGGCCCCCTCCGCTGATTGGCCCGTGGTGCAGGCAGCA 653

Qy         64 GCCCGGCAGGCACGCTCCTGGTC-CGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 122
              |||||||||||||||||||||| |  ||||||||||||||||||||||||||||||||||
Db        654 GCCCGGCAGGCACGCTCCTGGCCGGGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACG 713

Qy        123 ATT----AGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCA-CAGAGGCCGAGGACCC 177
                | |     |||||||||||| |||||||||||||| ||||||| ||| || |||| |||
```

```
Db        714 ACTTGCATGCAGCTCAGAAATCCCTCTGGGTCTCATCACTGCAGCAGTGGTCGAGTACCT 773

Qy        178 CCTCCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCT 237

Db        774 CCTCGGAGCTTTTCTACGACTTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGTT 833

Qy        238 CAGCAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACC- 296

Db        834 TAGCAGAACTTCAGAGGGAGCAGAGAGGCTCAGCTATCCACTGCTGCTTGACACTGACCC 893

Qy        297 -------------------------ACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTA 330

Db        894 TATCCACTGCTGCTTGTCACTGACTGACCTGCTGCTCTCTATTCTTTTGAGTCGGGAG-A 952

Qy        331 ACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGGAGTGGGT----------GG 380

Db        953 ACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGGCGCGCGGGGTGTGTGTGGG 1012

Qy        381 GCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTG--TCACACCCCCCTTC 438

Db       1013 GGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGGCTGGCACACACACACCTTC 1072

Qy        439 CATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGT 498

Db       1073 CATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGT 1132

Qy        499 CCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 558

Db       1133 CCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATT 1192

Qy        559 CCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACT 618

Db       1193 CCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACT 1252

Qy        619 GCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCG 678

Db       1253 GCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCG 1312

Qy        679 AGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCA 738

Db       1313 AGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCA 1372

Qy        739 ACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 798

Db       1373 ATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCC 1432

Qy        799 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGG 858

Db       1433 CGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGG 1492

Qy        859 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGC 918

Db       1493 CACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGC 1552

Qy        919 CCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACT 978

Db       1553 CCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACT 1612

Qy        979 CCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCC 1038

Db       1613 CCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCC 1672

Qy       1039 TGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCT 1098

Db       1673 TGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCT 1732

Qy       1099 TGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAG 1158

Db       1733 TGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTGGAAAGGGAGGGGACCAGAG 1792

Qy       1159 CTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGCCCCTTCTGGCTTTCATTA 1218

Db       1793 CCGTCTGGAGTGGGAGGTAGTGGAGGCTCTCAAGCATCTCGCCTCTTCTGGCTTTCACTA 1852

Qy       1219 GTCAGGTCC 1227

Db       1853 CTTGGATCC 1861


RESULT 2
US-10-004-717-24
; Sequence 24, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 861
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-24

  Query Match            46.7%;  Score 682.4;  DB 14;  Length 861;
  Best Local Similarity  89.7%;  Pred. No. 1.5e-185;
  Matches  771;  Conservative    0;  Mismatches   76;  Indels   13;  Gaps    3;

Qy        312 TTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAAGAGGGG 371

Db          2 TTCTTTTGAGTCGGGAG-AACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGG 60

Qy        372 AGTGGGT----------GGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGA 421

Db         61 CGCGCGGGGTGTGTGTGGGGGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGG 120

Qy        422 CTG--TCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGAT 479

Db        121 CTGGCACACACACACCTTCCATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGAT 180

Qy        480 GCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCAC 539

Db        181 GCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCAC 240

Qy        540 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCC 599

Db        241 GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCC 300

Qy        600 GAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGG 659

Db        301 GAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGG 360

Qy        660 CGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAG 719

Db        361 CGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAG 420

Qy        720 GCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGC 779

Db        421 GCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGC 480

Qy        780 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 839

Db        481 GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC 540
```

```
Qy        840 GCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAC 899
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAT 600

Qy        900 GGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGC 959
              ||||| |||||||||||||||||||| |||||||| ||||| ||||| ||||||| |||
Db        601 GGCCCGGAGCCCCCTGTGCCCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGG 660

Qy        960 GACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCA 1019
              |||||||||||||||||||||||| |||||||| |||| |||||||||||| ||||||
Db        661 GACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCA 720

Qy       1020 TTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACC 1079
              |||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||| |
Db        721 TTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCA 780

Qy       1080 CTGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCA 1139
              |||||||||||||||||||||||| | ||    | ||| ||||| |||||| ||| |
Db        781 CTGGTGTTCTCAGACTTCTTGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTG 840

Qy       1140 GAAAGGGAGGGAGTCAGAGC 1159
              |||||||||||   || |||
Db        841 GAAAGGGAGGGGACCACAGC 860
```

RESULT 3
US-10-004-717-4
; Sequence 4, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-4

Query Match 39.9%; Score 582.6; DB 14; Length 645;
Best Local Similarity 94.0%; Pred. No. 5.9e-157;
Matches 606; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```
Qy        459 ATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCC 518
              |||||||||||||||||||||||| |||||||||||||||| |||||| || |||||
Db          1 ATGGCGCCTCATCCCTTGGATGCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCT 60

Qy        519 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 578
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACT 120

Qy        579 CTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAG 638
              ||| |||| |||||||||||||||||||| |||||||||||||||| |||||||||
Db        121 CTCATACCTAGGGACTGCTCCGAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAG 180

Qy        639 CTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAG 698
              ||||| || || ||||||||||||||||||||||||||||||||||||||| ||||| |||
Db        181 CTCCGCGCCCGACGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAG 240

Qy        699 CGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAAC 758
              ||| |||||||||||||||||||||| || |||||||||| |||||||||||||| |||
Db        241 CGAAGAAGCCGGCGCAAGAAGGCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAAC 300

Qy        759 TCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 818
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCGGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACA 360

Qy        819 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGC 878
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        361 AAGATCGAGACCCTGCGCTTCGCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGC 420

Qy        879 ATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGC 938
              |||||||||||||||||| ||||| |||||||||||||||||| |||||||| |||
Db        421 ATAGCGGACCACAGCTTCTATGGCCCGGAGCCCCCTGTGCCCTGTGGAGAGCTGGGGAGC 480

Qy        939 CCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGT 998
              || ||||| ||||||| ||| ||||||||||||||||||||||||| |||||||| |||
Db        481 CCCGGAGGTGGCTCCAACGGGGACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCGGGT 540

Qy        999 AGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCA 1058
              | |||||||||||| |||||||||||||| ||||||||||||||||||||||||||||||
Db        541 AACCTGAGCCCCACGGCCTCATTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCCCCA 600

Qy       1059 TCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTGTGA 1103
              |||| ||||||||||||  | |||||||||||||||||||||||
Db        601 TCCTATCTGCTCCCGGGAGCACTGGTGTTCTCAGACTTCTTGTGA 645
```

RESULT 4
US-09-817-360-1
; Sequence 1, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5340
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-817-360-1

Query Match 35.4%; Score 517.4; DB 9; Length 5340;
Best Local Similarity 67.4%; Pred. No. 4.5e-138;
Matches 826; Conservative 0; Mismatches 351; Indels 48; Gaps 5;

```
Qy         20 GTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCAGCAGCCCGGCAGGCACGCT 79
              | |||  || |||||  |||||||||| ||||| | |||||||||| ||||||||||
Db       2535 GGCCCCAGGGCCCCGGCGCTGATTGGCCGGTGGCGCGGGCAGCAGCCGGGCAGGCACGCT 2594

Qy         80 CCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACACGATTAGCAGCTCAGAAGT 139
              ||||| ||||||  |||||||||||||||| ||| ||||||| | || ||||||
Db       2595 CCTGGCCCGGGCGAAGCAGATAAAGCGTGCCAAGGGGCACACGACTTGCTGCTCAGGAAA 2654

Qy        140 CCCTCTGGGTCTCACCACTGC----------------ACAGAGGCCGAGGACCCCCTC 181
              ||    ||||||||| | ||                  ||||||||  ||||||| |
Db       2655 TCCCTGCGGTCTCACCGCCGCGCCTCGAGAGAGAGCGTGACAGAGGCCTCGGACCCCATT 2714

Qy        182 CGAGCTTCTTTGCTGCCTCCAGACGCAATTTACT-CCAGGCGAGGGCGCCTGCAGCTCAG 240
              |   ||||||| || | |   | |       ||| ||||||| ||||||||||||||||
Db       2715 CTCTCTTCTTTTCTCCTTTGGGGCTGGGGCAACTCCCAGGCGGGGGCGCCTGCAGCTCAG 2774

Qy        241 CAAAACTTCGAAGCGAGCAG----------AGGGGTTCAGCTATCCACCGCTGCT 285
```

```
Db       2775 CTGAACTTGGCGACCAGAAGCCCGCTGAGCTCCCCACGGCCCTCGCTGCTCATCGCTCTC 2834

Qy        286 TGACTCTGACCACCCGCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTA 345

Db       2835 TATTCTTTTGCGCCGGTAGAAAGGTAATATTTGGAGGCCTCCGAGGGACGGGCAGGGGAA 2894

Qy        346 GGAAC-CTCCAAAGGGTAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGT 404

Db       2895 AGAGGGATCCTCTGACCCAGCGGGGGCTGGGAGGATGGCTGTTTTTGTTTTTCCCACCT 2954

Qy        405 GACCTCTAAGTCAGAGACTG------------TCACACCCCCCTTCCATTTTTTCCCAA 451

Db       2955 AGCCTCGGAATCGCGGACTGCGCCGTGACGGACTCAAACTTACCCTTCCCTCTGACCCCG 3014

Qy        452 CCTCAGGATGGCGCCTCATCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCA 511

Db       3015 CCGTAGGATGACGCCTCAACCCTCGGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGA 3074

Qy        512 GCAACCCTTTCCCGGAGCCTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAG 571

Db       3075 GCGGTCCTTCCCCAGAGCCTCGGAAGACGAAGTGACCTGCCCCACGTCCGCCCCGCCCAG 3134

Qy        572 CCCCACTCTCGTACCGAGGGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATC 631

Db       3135 CCCCACTCGCACACGGGGGAACTGCGCAGAGGCGGAAGAGGGAGGCTGCCGAGGGGCCCC 3194

Qy        632 GAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAG 691

Db       3195 GAGGAAGCTCCGGGCACGGCGCGGGGGACGCAGCCGGCCTAAGAGCGAGTTGGCACTGAG 3254

Qy        692 CAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAA 751

Db       3255 CAAGCAGCGACGGAGTCGGCGAAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACAA 3314

Qy        752 CCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAA 811

Db       3315 CCTCAACTCGGCACTGGACGCCCTGCGCGGTGTCCTGCCCACCTTCCCAGACGACGCGAA 3374

Qy        812 ACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGAC 871

Db       3375 GCTCACCAAGATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTGACTCAAAC 3434

Qy        872 GCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCT 931

Db       3435 GCTGCGCATAGCGGACCACAGCTTGTACGCGCTGGAGCCGCCGGCGCCGCACTGCGGGGA 3494

Qy        932 GGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCA 991

Db       3495 GCTGGGCAGCCCAGGCGGTTCCCCCGGGGACTGGGGGTCCCTCTACTCCCCAGTCTCCCA 3554

Qy        992 AGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAG 1051

Db       3555 GGCTGGCAGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCGGGCTGCTGGGGGCCAC 3614

Qy       1052 CTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTGTTCTCAGACTTCTTGTGAAGGGCCCA 1111

Db       3615 CTCTTCCGCCTGCTTGAGCCCAGGCAGTCTGGCTTTCTCAGATTTTCTGTGAAAGGACCT 3674

Qy       1112 AACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGGGAGGGAGTCAGAGCTGTCTGAAATGG 1171

Db       3675 GTCTGTCGCTGGGCTGTGGGTGCTAAGGGTAAGGGAGAGGGAGGGAGCCGGGAGCCGTAG 3734

Qy       1172 AAGGTAGTGGAGGCACTCGAGCATC 1196

Db       3735 AGGGTGGCCGACGGCGGCGGCCCTC 3759


RESULT 5
US-08-722-570-12
; Sequence 12, Application US/08722570
; Publication No. US20030044887A1
;  GENERAL INFORMATION:
;     APPLICANT:  Anderson, David J.
;     APPLICANT:  Ma, Qiufu
;     TITLE OF INVENTION:  NEUROGENIN
;     NUMBER OF SEQUENCES:  20
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;       STREET:  Four Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  United States
;       ZIP:  94111-4187
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/722,570
;       FILING DATE:  27-SEP-1996
;       CLASSIFICATION:  5365
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Silva, Robin M.
;       REGISTRATION NUMBER:  38,304
;       REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-3249
;       TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  12:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1527 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA
US-08-722-570-12

  Query Match             9.9%;  Score 145.2;  DB 8;  Length 1527;
  Best Local Similarity   69.2%;  Pred. No. 2e-31;
  Matches  198;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        635 GAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAA 694

Db        456 GCAGGAGCGGCGGCGACGGCGAGGTCGCGCGCGGGTGCGGTCCGAGGCGCTGCTGCACTC 515

Qy        695 GCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCT 754

Db        516 GCTGCGGAGGAGCCGTCGCGTCAAGGCCAACGATCGCGAGCGCAACCGTATGCATAACCT 575

Qy        755 TAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACT 814

Db        576 CAACGCTGCGCTGGACGCTCTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCT 635

Qy        815 TACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCT 874

Db        636 CACCAAGATTGAGACGCTGCGCTTCGCCTACAACTACATCTGGGCCCTGGCTGAGACACT 695

Qy        875 GCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC 920

Db        696 GCGCCTGGCAGATCAAGGGCTCCCGGGGGGCGGTGCCCGGGAGCGC 741


RESULT 6
US-08-722-570-13
; Sequence 13, Application US/08722570
; Publication No. US20030044887A1
;  GENERAL INFORMATION:
;     APPLICANT:  Anderson, David J.
;     APPLICANT:  Ma, Qiufu
;     TITLE OF INVENTION:  NEUROGENIN
;     NUMBER OF SEQUENCES:  20
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
```

```
;       STREET:  Four Embarcadero Center, Suite 3400
;       CITY:  San Francisco
;       STATE:  California
;       COUNTRY:  United States
;       ZIP:  94111-4187
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/722,570
;       FILING DATE:  27-SEP-1996
;       CLASSIFICATION:  5365
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Silva, Robin M.
;       REGISTRATION NUMBER:  38,304
;       REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 781-1989
;       TELEFAX:  (415) 398-3249
;       TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  738 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA
US-08-722-570-13

  Query Match             9.6%;  Score 140.4;  DB 8;  Length 738;
  Best Local Similarity   68.2%;  Pred. No. 4.1e-30;
  Matches  195;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy        635 GAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAA 694
              | ||   ||   | ||||  |||| |||    ||     |  ||||      |||  |
Db        207 GCAGGAACGGCGGAGGCGGCGAGGTCGCGCTCGGGTGCGGTCCGAGGCTCTGCTGCACTC 266

Qy        695 GCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCT 754
              | |||  | || || |||   || |||||||| || ||||||||||||||||||||||
Db        267 CCTGCGGAGGAGTCGTCGCGTCAAAGCCAACGATCGCGAGCGCAACCGCATGCACAACCT 326

Qy        755 TAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACT 814
              ||| | ||||||||  ||  ||||| | || ||||||  | |||||  || ||| |||| ||
Db        327 CAACGCTGCGCTGGACGCCTTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCT 386

Qy        815 TACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCT 874
              || ||||| ||||| ||||||||||| |||||||||| ||||| ||| || |||| ||
Db        387 CACCAAGATTGAGACGCTGCGCTTCGCCTACAACTACATCTGGGCCCTGGCTGAGACACT 446

Qy        875 GCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCCC 920
              |||| | || || ||  |  ||  |||  |    |||  | || |
Db        447 GCGCCTGGCAGATCAAGGGCTCCCCGGGGGCAGTGCCCGGGAGCGC 492


RESULT 7
US-10-004-717-30
; Sequence 30, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
;   LENGTH: 1385
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-30

  Query Match             9.6%;  Score 139.8;  DB 14;  Length 1385;
  Best Local Similarity   78.1%;  Pred. No. 7e-30;
  Matches  168;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
              | |||  |    ||||||||| |||| |||||||||||||||||||||| ||| |||||
Db        715 ACCCGCAGGCTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCG 774

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
              ||||| ||||||||||  || |||||||||||||| || || ||||| || || ||||||
Db        775 CTGGACGCGCTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATC 834

Qy        825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCG 884
              ||||| |||||||||||||||| ||||| ||||| || ||  |||| |||||| | |||
Db        835 GAGACGCTGCGCTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCG 894

Qy        885 GACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCC 919
              |||||| ||  |  || |   |  | ||  | | |
Db        895 GACCACTGCGCCGGCGCCGGTGGCCTCCAGGGGGC 929


RESULT 8
US-10-004-717-6
; Sequence 6, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 1412
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-6

  Query Match             9.6%;  Score 139.8;  DB 14;  Length 1412;
  Best Local Similarity   78.1%;  Pred. No. 7e-30;
  Matches  168;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
              | |||  |    ||||||||| |||| |||||||||||||||||||||| ||| |||||
Db        418 ACCCGCAGGCTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCG 477

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
              ||||| ||||||||||  || |||||||||||||| || || ||||| || || ||||||
Db        478 CTGGACGCGCTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATC 537

Qy        825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCG 884
              ||||| |||||||||||||||| ||||| ||||| || ||  |||| |||||| | |||
Db        538 GAGACGCTGCGCTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCG 597
```

```
Qy        885 GACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCC 919
              |||||| ||  |  || |   |  | ||  | | |
Db        598 GACCACTGCGCCGGCGCCGGTGGCCTCCAGGGGGC 632


RESULT 9
US-10-004-717-37
; Sequence 37, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
;   LENGTH: 1412
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-37

  Query Match             9.6%;  Score 139.8;  DB 14;  Length 1412;
  Best Local Similarity   78.1%;  Pred. No. 7e-30;
  Matches  168;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
              | |||  |    |||||||| |||| ||||||||||||||||||||||| ||| |||||
Db        418 ACCCGCAGGCTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCG 477

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
              ||||| ||||||||||  || |||||||||||||| || || ||||| || || ||||||
Db        478 CTGGACGCGCTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATC 537

Qy        825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCG 884
              ||||| |||||||||||||||| ||||| ||||| || ||  |||| |||||| | |||
Db        538 GAGACGCTGCGCTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCG 597

Qy        885 GACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCC 919
              |||||| ||  |  || |   |  | ||  | | |
Db        598 GACCACTGCGCCGGCGCCGGTGGCCTCCAGGGGGC 632


RESULT 10
US-10-004-717-20
; Sequence 20, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
;   LENGTH: 790
;   TYPE: DNA
;   ORGANISM: chicken
US-10-004-717-20

  Query Match             9.5%;  Score 139.2;  DB 14;  Length 790;
  Best Local Similarity   84.8%;  Pred. No. 9.3e-30;
  Matches  156;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy        702 CGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCC 761
              || ||||||||   ||| |||||| |||| |||||||||||||||||||||||| ||| |
Db        344 CGGAGCCGGCGGCTGAAAGCCAACAACCGCGAGCGCAACCGCATGCACAACCTGAACGCG 403

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              |||||||| |||||||||| || |||||||||||||| || |||||||| || || |||
Db        404 GCGCTGGACGCGCTGCGCGACGTGCTGCCCACCTTCCCCGAGGACGCCAAGCTCACCAAG 463

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATA 881
              |||||||| |||||||||||||||||||||| ||||| || ||  |||||||||| |
Db        464 ATCGAGACGCTGCGCTTCGCCCACAACTACATCTGGGCGCTCACCGAGACGCTGCGCCTG 523

Qy        882 GCGG 885
              || |
Db        524 GCCG 527


RESULT 11
US-10-004-717-18
; Sequence 18, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 1074
;   TYPE: DNA
;   ORGANISM: chicken
US-10-004-717-18

  Query Match             9.5%;  Score 138.2;  DB 14;  Length 1074;
  Best Local Similarity   71.4%;  Pred. No. 1.9e-29;
  Matches  182;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy        642 CGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGA 701
              ||   | || |  | || ||     || |   || ||||      ||| ||  |   |
Db        372 CGGCGGAGGAGACGCGGCCGTGCGCGGGCGCGGACCGAGGCTTTGCTGCACACCCTCAAA 431

Qy        702 CGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCC 761
              || ||||| ||   ||| |||||||||||||||| |||||||||||| |||| ||| ||
Db        432 CGGAGCCGCCGGGTGAAAGCCAACGACCGGGAGCGGAACCGCATGCACCACCTCAACGCC 491

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              |||||||||| ||| ||| | |||||||| |||||||| || ||| ||||||| || ||
Db        492 GCGCTGGATGAGCTCCGCAGCGTCCTGCCGACCTTCCCCGACGACACCAAACTCACCAAA 551

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATA 881
```

```
                |||||  ||||||||||||||||  |||||||||| |||||  ||  |  ||||  ||  ||  |
Db        552 ATCGAAACCCTGCGCTTCGCTTACAACTACATCTGGGCCCTCTCCGAGACCCTTCGTTTG 611

Qy        882 GCGGACCACAGCTTC 896
              ||  ||  ||   ||  ||
Db        612 GCCGAGCAGTGCCTC 626


RESULT 12
US-08-722-570-14
; Sequence 14, Application US/08722570
; Publication No. US20030044887A1
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  5365
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1312 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-14

  Query Match             7.9%;  Score 115.6;  DB 8;  Length 1312;
  Best Local Similarity   67.4%;  Pred. No. 6e-23;
  Matches  163;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy        647 GCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAG 706
              | || || |||| |||      |  ||||||||      |  ||   |||          |
Db        539 GAGGAGCCGAGGCCGCGCTCAGGGCAAGAGCGGAGAAACTGTGTTAAAGATCAAGAAGAC 598

Qy        707 CCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCT 766
              |||||||   || || ||| |||||||  | || ||||||||||||| |||||  |||||
Db        599 CCGGCGCGTTAAAGCTAACAACCGGGAAAGGAATCGCATGCACAACCTGAACTCTGCGCT 658

Qy        767 GGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGA 826
              |||  |  ||  |  |  ||  |||||  |  || || || || ||||||||  || |||||  ||
Db        659 TGATTCCCTCAGGGAAGTGTTGCCCTCTTTACCTGAAGATGCCAAACTCACCAAGATAGA 718

Qy        827 GACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGA 886
              ||||  |||||||  |||  ||||||||||  |||||  ||  |    |  ||   |||||  |  |  ||
Db        719 GACCTTGCGCTTTGCCTACAACTACATCTGGGCTCTTAGCGAAACTTTGCGCCTTGGCGA 778

Qy        887 CC 888
              ||
Db        779 CC 780


RESULT 13
US-08-722-570-15
; Sequence 15, Application US/08722570
; Publication No. US20030044887A1
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  5365
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1277 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-15

  Query Match             7.5%;  Score 109.8;  DB 8;  Length 1277;
  Best Local Similarity   74.6%;  Pred. No. 2.8e-21;
  Matches  138;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
              | |||||||   || |||||  |||| ||| | || |||||||||  |||| ||||  |||
Db        439 ACCCGGCGCGTTAAAGCCAATAACCGCGAGAGGAATCGCATGCACCACCTGAACTATGCG 498

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
              || ||| | ||| | |  || || ||  | || || || ||||||||||| || |||||
Db        499 CTCGATTCTCTGAGGGAGGTTCTACCGTCATTACCCGAAGACGCCAAACTCACCAAGATA 558

Qy        825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCG 884
              |||||| ||||||| |||||||||||||| |||||  || |   | ||  |||||  | ||
Db        559 GAGACCTTGCGCTTTGCCCACAACTACATCTGGGCTCTTAGCGAAACTTTGCGCCTGGCC 618

Qy        885 GACCA 889
              |||||
Db        619 GACCA 623


RESULT 14
US-10-413-358-27
; Sequence 27, Application US/10413358
; Publication No. US20030219894A1
; GENERAL INFORMATION:
;  APPLICANT: Susumu Seino, JCR Pharmaceuticals Co., Ltd.
;  TITLE OF INVENTION: Induction of insulin-producing cells
;  FILE REFERENCE: GP55
```

```
;  CURRENT APPLICATION NUMBER: US/10/413,358
;  CURRENT FILING DATE:  2003-04-15
;  PRIOR APPLICATION NUMBER: JP2002-115064
;  PRIOR FILING DATE: 2002-04-17
;  NUMBER OF SEQ ID NOS: 27
; SEQ ID NO 27
;   LENGTH: 1099
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-413-358-27

  Query Match             6.1%;  Score 89;  DB 13;  Length 1099;
  Best Local Similarity   61.4%;  Pred. No. 2.5e-15;
  Matches  143;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy        650 GCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCG 709
              | |  | || | ||| | |   ||||       ||| |      |||       |    |
Db        258 GAGACGCGGCCCCAAAAAGAAGAAGATGACTAAGGCTCGCCTGGAGCGTTTTAAATTGAG 317

Qy        710 GCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGA 769
              |||| ||||||  |||| |||||||||  ||||||||||||   || ||| | ||||| ||
Db        318 ACGCATGAAGGCTAACGCCCGGGAGCGGAACCGCATGCACGGACTGAACGCGGCGCTAGA 377

Qy        770 TGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGAC 829
              ||||||   ||  ||||  ||   |  |       || ||  | || ||||||||
Db        378 CAACCTGCGCAAGGTGGTGCCTTGCTATTCTAAGACGCAGAAGCTGTCCAAAATCGAGAC 437

Qy        830 CCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAG 882
              |||||||| ||| | |||||||| ||||| ||| |  |||  ||||||  ||
Db        438 TCTGCGCTTGGCCAAGAACTACATCTGGGCTCTGTCGGAGATCCTGCGCTCAG 490


RESULT 15
US-10-413-358-26
; Sequence 26, Application US/10413358
; Publication No. US20030219894A1
; GENERAL INFORMATION:
;  APPLICANT: Susumu Seino, JCR Pharmaceuticals Co., Ltd.
;  TITLE OF INVENTION: Induction of insulin-producing cells
;  FILE REFERENCE: GP55
;  CURRENT APPLICATION NUMBER: US/10/413,358
;  CURRENT FILING DATE:  2003-04-15
;  PRIOR APPLICATION NUMBER: JP2002-115064
;  PRIOR FILING DATE: 2002-04-17
;  NUMBER OF SEQ ID NOS: 27
; SEQ ID NO 26
;   LENGTH: 1211
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-413-358-26

  Query Match             6.1%;  Score 89;  DB 13;  Length 1211;
  Best Local Similarity   61.4%;  Pred. No. 2.5e-15;
  Matches  143;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy        650 GCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCG 709
              | |  | || | ||| | |   ||||       ||| |      |||       |    |
Db        348 GAGACGCGGCCCCAAAAAGAAGAAGATGACTAAGGCTCGCCTGGAGCGTTTTAAATTGAG 407

Qy        710 GCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGA 769
              |||| ||||||  |||| |||||||||  ||||||||||||   || ||| | ||||| ||
Db        408 ACGCATGAAGGCTAACGCCCGGGAGCGGAACCGCATGCACGGACTGAACGCGGCGCTAGA 467

Qy        770 TGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGAC 829
              ||||||   ||  ||||  ||   |  |       || ||  | || ||||||||
Db        468 CAACCTGCGCAAGGTGGTGCCTTGCTATTCTAAGACGCAGAAGCTGTCCAAAATCGAGAC 527

Qy        830 CCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAG 882
              |||||||| ||| | |||||||| ||||| ||| |  |||  ||||||  ||
Db        528 TCTGCGCTTGGCCAAGAACTACATCTGGGCTCTGTCGGAGATCCTGCGCTCAG 580


Search completed: January 26, 2004, 22:15:27

Job time : 520 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 26, 2004, 18:46:56 ; Search time 3301 Seconds
(without alignments)
10749.632 Million cell updates/sec

Title: US-09-595-947E-1
Perfect score: 1460
Sequence: 1 gcaggtagcgagaggagcag.....agagtgacctaatccagtgt 1460

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	763.2	52.3	1540	11	AK008017	AK008017 Mus muscu
	2	545	37.3	1025	14	BY708009	BY708009 BY708009
c	3	473.4	32.4	593	28	AZ296526	AZ296526 RPCI-23-1
	4	133.4	9.1	687	14	CA945402	CA945402 UI-M-FD0-
	5	133	9.1	1001	9	AL540071	AL540071 AL540071
	6	130.2	8.9	632	12	BJ010277	BJ010277 BJ010277
	7	128.6	8.8	600	12	BG808248	BG808248 2083-52 M
	8	128.6	8.8	730	13	BU612495	BU612495 UI-M-FR0-
	9	127.4	8.7	823	13	BU054481	BU054481 UI-M-FD0-
	10	121.8	8.3	947	13	BX419330	BX419330 BX419330
	11	120.4	8.2	600	13	BU924937	BU924937 7103-91 M
	12	118.6	8.1	932	14	CA979119	CA979119 AGENCOURT
c	13	115.6	7.9	1037	29	CNS03VJ9	AL262494 Tetraodon
	14	110.8	7.6	401	9	AW147434	AW147434 da02h12.y
	15	110.2	7.5	603	14	CD282259	CD282259 G39271.16
	16	110.2	7.5	865	14	CA496091	CA496091 AGENCOURT
	17	110.2	7.5	1098	13	BQ169355	BQ169355 STR00303
	18	107.4	7.4	1039	10	BE780690	BE780690 601469349
	19	105.2	7.2	604	13	BX308104	BX308104 BX308104
	20	104.6	7.2	814	13	BQ178789	BQ178789 UI-M-EV0-
	21	104	7.1	629	12	BJ093114	BJ093114 BJ093114
	22	103.6	7.1	595	12	BJ030202	BJ030202 BJ030202
	23	96.4	6.6	704	13	BU057851	BU057851 UI-M-FR0-
	24	96.4	6.6	740	14	CA319439	CA319439 UI-M-FW0-
	25	96.4	6.6	781	13	BU611678	BU611678 UI-M-FI0-
	26	96.4	6.6	835	14	CA320553	CA320553 UI-M-FW0-
	27	96.4	6.6	953	9	AU067624	AU067624 AU067624
	28	93.8	6.4	889	12	BI734232	BI734232 603351339
	29	92.4	6.3	600	12	BG802832	BG802832 0191-69 M
	30	92.4	6.3	600	12	BG804072	BG804072 0251-06 M
	31	92.4	6.3	651	10	BG296688	BG296688 602393635
	32	92.4	6.3	658	14	BY719254	BY719254 BY719254
	33	92.4	6.3	678	10	BG296579	BG296579 602394490
	34	92.4	6.3	692	12	BI737697	BI737697 603358625
	35	92.4	6.3	752	12	BI732893	BI732893 603355021
	36	92.4	6.3	757	12	BI733773	BI733773 603352396
	37	92.4	6.3	782	12	BI732663	BI732663 603354913
	38	92.4	6.3	792	14	CB249320	CB249320 UI-M-FC0-
	39	92.4	6.3	835	10	BG298117	BG298117 602396478
	40	92.4	6.3	836	13	BU604426	BU604426 AGENCOURT
	41	92.4	6.3	936	13	BU506118	BU506118 AGENCOURT
	42	92.4	6.3	951	12	BI731777	BI731777 603353558
	43	92.4	6.3	951	13	BQ891444	BQ891444 AGENCOURT
	44	92.4	6.3	968	12	BI731880	BI731880 603353625
	45	92.4	6.3	994	10	BB610830	BB610830 BB610830

ALIGNMENTS

RESULT 1
AK008017
LOCUS AK008017 1540 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010001M19 product:atonal homolog 5 (Drosophila), full insert sequence.
ACCESSION AK008017
VERSION AK008017.1 GI:12841941
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1540)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

```
FEATURES             Location/Qualifiers
     source          1. .1540
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:2010001M19"
                     /db_xref="MGI:1907403"
                     /db_xref="taxon:10090"
                     /clone="2010001M19"
                     /sex="male"
                     /tissue_type="small intestine"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             241. .885
                     /note="unnamed protein product; atonal homolog 5
                     (Drosophila) (MGD|MGI:893591, GB|NM_009719, evidence:
                     BLASTN, 99%, match=648)
                     putative"
                     /codon_start=1
                     /protein_id="BAB25411.1"
                     /db_xref="GI:12841942"
                     /db_xref="MGI:893591"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSHTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKKRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKVETLRFAHNYIWALTQTLRIADHSLYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT      306 a    485 c    397 g    352 t
ORIGIN
```

```
  Query Match           52.3%;  Score 763.2;  DB 11;  Length 1540;
  Best Local Similarity  77.7%;  Pred. No. 2.1e-176;
  Matches 1051;  Conservative    0;  Mismatches  183;  Indels  118;  Gaps    6;

Qy       127 GCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCA-CAGAGGCCGAGGACCCCCTCCGAG 185
            |||||||||||| |||||||||||||| ||||||| ||| || |||| ||| |||| |||
Db         4 GCAGCTCAGAAATCCCTCTGGGTCTCATCACTGCAGCAGTGGTCGAGTACCTCCTCGGAG 63

Qy       186 CTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAGCAAAA 245
            ||| | | |  | ||||||||||||||||||||||||||||||||||||| | |||| ||
Db        64 CTTTTCTACGACTTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGTTTAGCAGAA 123

Qy       246 CTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCCGCAGC 305
            ||||  || |||||||| || |||||||||| |||||||||| ||||||     ||
Db       124 CTTCAGAGGGAGCAGAGAGGCTCAGCTATCCACTGCTGCTTGACACTGACCCTATCCACT 183

Qy       306 TCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAA 365
              ||   |   |||       |   ||  ||  ||| |   |     ||
Db       184 GCTGCTTGTCACTGACTGACCTGCTGCTCTCTATTCTTTTGAGTCGGGAGAACT------ 237

Qy       366 GAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTGT 425

Db       238 ------------------------------------------------------------ 237

Qy       426 CACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCC 485
                                          |||||||||||||||||||||||||||||| |
Db       238 -----------------------------AGGATGGCGCCTCATCCCTTGGATGCGCTC 267

Qy       486 ACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTG 545
            |||||||||||||||| |||||| || ||||| ||||||||||||||||||||||||||||
Db       268 ACCATCCAAGTGTCCCCAGAGACACAACAACCTTTCCCGGAGCCTCGGACCACGAAGTG 327

Qy       546 CTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCA 605
            |||||||||||||||||||||||||||| ||||||| |||| ||||||||||||||||||
Db       328 CTCAGTTCCAATTCCACCCCACCTAGCCACACTCTCATACCTAGGGACTGCTCCGAAGCA 387

Qy       606 GAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAAC 665
```

```
Db        388 GAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGGCGCAAC 447

Qy        666 AGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAAC 725
              |||||||||||||||||||||| |||||  ||||| ||||||||||||||||||||||
Db        448 AGGCCCAAGAGCGAGTTGGCACTCAGCAAAAAGCGAAGAAGCCGGCGCAAGAAGGCCAAT 507

Qy        726 GACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTC 785
              || |||||||||||  |||||||||||||| ||||| ||||||||||||||||||||||
Db        508 GATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGCGGTGTC 567

Qy        786 CTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCAC 845
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        568 CTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGGTCGAGACCCTGCGCTTCGCCCAC 627

Qy        846 AACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCC 905
              ||||||| ||||||||||||||||||||||||||||||||||||||||| || |||||
Db        628 AACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTGTATGGCCCG 687

Qy        906 GAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGG 965
              |||||||||||||||||| |||||||| |||||  |||||  ||||||| ||| ||||||
Db        688 GAGCCCCCTGTGCCCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGGGACTGG 747

Qy        966 GGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAG 1025
              |||||||||||||||||| |||||||| |||| ||||||||||| ||||||||||||
Db        748 GGCTCTATCTACTCCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCATTGGAG 807

Qy       1026 GAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTG 1085
              || |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        808 GAATTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCACTGGTG 867

Qy       1086 TTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGG 1145
              |||||||||||||||||||| | ||    | ||| |||||  |||||| ||| |  ||||||
Db        868 TTCTCAGACTTCTTGTGAAGAGACCTGTCTGGCTCTGGGTGGTGGGTGCTAGTGGAAAGG 927

Qy       1146 GAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGCCCCTT 1205
              |||||   |||||| |||||  | ||| |||||||||||||  ||| ||||||||||  |||
Db        928 GAGGGGACCAGAGCCGTCTGGAGTGGGAGGTAGTGGAGGCTCTCAAGCATCTCGCCTCTT 987

Qy       1206 CTGGCTTTCATTAGTCAGGTCCCT----------GATTTAACCAGGATTCGCACAGT 1252
              |||||||||| || |  | |||||           |  ||||| ||| ||| ||  |
Db        988 CTGGCTTTCACTACTTGGATCCCTAGCCCTCTCACAGGGCTTAACTAGGCTTCTCATCGG 1047

Qy       1253 TCCTTGCTGCTGTGCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCA 1312
              | |   |      || | |||||   || ||||  ||||| ||| |||||||| ||||||
Db       1048 TAC---CCTTGCTGTGCGCACAGCAGACATTGGGGGCTGCTCTTCTCTTAACTCTCCTCG 1104

Qy       1313 GTGTGGCCACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGG 1372
              |||  ||||| |||||||| ||||||||||    ||||   ||||||||  ||||| ||
Db       1105 GTGCAGCCACATCAAACTCTCGCTCCAAGCATTTGAGAATGGTAGCACTACCTAGTT-GG 1163

Qy       1373 AGACTCCCATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCT----CCAACC 1428
              |||||||||||||||||||| || |||||| |||||||||| |||||    ||| ||
Db       1164 AGACTCCCATACTTCCTGGTGAGTCTGCCCTCATTCAAATCTGCCGGCCTCCGACCATCC 1223

Qy       1429 ACCGCTTTCTCCAGAGTGACCTAATCCAGTGT 1460
              | | |||| ||||| ||||||||||||||||||
Db       1224 ATCACTTTTTCCAGGGTGACCTAATCCAGTGT 1255
```

RESULT 2
BY708009
LOCUS BY708009 1025 bp mRNA linear EST 16-DEC-2002
DEFINITION BY708009 RIKEN full-length enriched, adult male small intestine Mus musculus cDNA clone 2010001M19 5', mRNA sequence.
ACCESSION BY708009
VERSION BY708009.1 GI:27119192
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1025)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .1025
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="2010001M19"
/sex="male"

/tissue_type="small intestine"
/dev_stage="adult"
/lab_host="SOLR"
/clone_lib="RIKEN full-length enriched, adult male small intestine"
/note="Site_1: XhoI; Site_2: SstI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI."

BASE COUNT 215 a 316 c 286 g 207 t 1 others
ORIGIN

Query Match 37.3%; Score 545; DB 14; Length 1025;
Best Local Similarity 77.4%; Pred. No. 6.1e-123;
Matches 820; Conservative 0; Mismatches 135; Indels 104; Gaps 9;

```
Qy        127 GCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCA-CAGAGGCCGAGGACCCCCTCCGAG 185
              |||||||||| |||||||||||||| ||||||| ||| || |||| ||| |||| |||
Db          4 GCAGCTCAGAAATCCCTCTGGGTCTCATCACTGCAGCAGTGGTCGAGTACCTCCTCGGAG 63

Qy        186 CTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAGCAAAA 245
              ||| | | |  | ||||||||||||||||||||||||||||||||||||| | |||| ||
Db         64 CTTTTCTACGACTTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGTTTAGCAGAA 123

Qy        246 CTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCCGCAGC 305
              ||||  || |||||||| || |||||||||||| |||||||||| ||||||     ||
Db        124 CTTCAGAGGGAGCAGAGAGGCTCAGCTATCCACTGCTGCTTGACACTGACCCTATCCACT 183

Qy        306 TCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGGTAGAA 365
              ||   |     |||        |   ||   ||   ||| |   |      ||
Db        184 GCTGCTTGTCACTGACTGACCTGCTGCTCTCTATTCTTTTGAGTCGGGAGAACT------ 237

Qy        366 GAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAGACTGT 425

Db        238 ------------------------------------------------------------ 237

Qy        426 CACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCC 485
                                          |||||||||||||||||||||||||||| |
Db        238 ----------------------------AGGATGGCGCCTCATCCCTTGGATGCGCTC 267

Qy        486 ACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTG 545
              ||||||||||||||| |||||| || ||||| ||||||||||||||||||||||||||||
Db        268 ACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCACGAAGTG 327

Qy        546 CTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCA 605
              |||||||||||||||||||||||||| ||||||| |||| ||||||||||||||||||
Db        328 CTCAGTTCCAATTCCACCCCACCTAGCCACACTCTCATACCTAGGGACTGCTCCGAAGCA 387

Qy        606 GAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAAC 665
              ||||  |||||||||||||||| |||||||||||||| || || ||||||||||||||||
Db        388 GAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGGCGCAAC 447

Qy        666 AGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAAC 725
              |||||||||||||||||||||||| |||||  ||||| ||||||||||||||||||||||
Db        448 AGGCCCAAGAGCGAGTTGGCACTCAGCAAAAAGCGAAGAAGCCGGCGCAAGAAGGCCAAT 507

Qy        726 GACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTC 785
              || |||||||||| |||||||||||||| ||||| ||||||||||||||||||||||||
Db        508 GATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGCGGTGTC 567

Qy        786 CTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCAC 845
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
```

```
Db        568 CTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGGTCGAGACCCTGCGCTTCGCCCAC 627

Qy        846 AACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCC 905
              |||||||| ||||||||||||||||||||||||||||||||||||||||| || |||||
Db        628 AACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTGTATGGCCCG 687

Qy        906 GAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGG 965
              |||||||||||||||||||| |||||||| | ||| ||| | ||||||| ||| ||||||
Db        688 GAGCCCCCTGTGCCCTGTGGAGAGCTGGGGA-CCCCGGACGTGGCTCCAACGGGGACTGG 746

Qy        966 GGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCATTGGAG 1025
              |||||||||||||| |||| |||||||| |||| |||||| ||||| ||||||||||||
Db        747 GGCTCTATCTACTCCTCAGTCTCCCAAGCGGGTAACCTGAG-CCCACGGCCTCATTGGAG 805

Qy       1026 GAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCCTGGTG 1085
              || || |||||||||||||| |||||||||||| || ||||||||||| | ||||||
Db        806 GA-TTACCTGGCCTGCAGGTG-CCAGCTCCCCAT-CTATCTGCTCCCGGGAGCACTGGTG 862

Qy       1086 TTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAGAAAGG 1145
              |||||||||||||||||||| |  |  | ||| ||||| |||||| ||| |    ||||
Db        863 TTCTCAGACTTCTTGTGAAGAG-ACTGTCTGGCTCTGGGTGGTGGGTGCTAGTGAGAAGG 921

Qy       1146 GAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGG 1184
              ||||| | ||   | ||||| |   | ||||||||||||
Db        922 GAGGG-GACACGACCGTCTGGACGTGGAGGTAGTGGAGG 959
```

RESULT 3
AZ296526/c
LOCUS AZ296526 593 bp DNA linear GSS 27-JUL-2000
DEFINITION RPCI-23-160G18.TV RPCI-23 Mus musculus genomic clone RPCI-23-160G18, genomic survey sequence.
ACCESSION AZ296526
VERSION AZ296526.1 GI:9538311
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 593)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished
COMMENT Other_GSSs: RPCI-23-160G18.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 160 row: G column: 18
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .593
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-160G18"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:

EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "

BASE COUNT 88 a 159 c 213 g 133 t
ORIGIN

Query Match 32.4%; Score 473.4; DB 28; Length 593;
Best Local Similarity 91.5%; Pred. No. 1.8e-105;
Matches 535; Conservative 0; Mismatches 46; Indels 4; Gaps 3;

```
Qy      426 CACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATGCGCCC 485
            ||||| | |||||||||||||||||||| ||| | |||||||||||||||||||||| |
Db      586 CACACACACCTTCCATTTTTTCCCAACCGCAGAAGGGCGCCTCATCCCTTGGATGCGCTC 527

Qy      486 ACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACGAAGTG 545
            |||||||||||||||| |||||| || ||||| ||||||||||||||||||||||||||
Db      526 ACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCACGAAGTG 467

Qy      546 CTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCGAAGCA 605
            ||||||||||||||||||||||||||||||||||| |||| ||||||||||||||||||
Db      466 CTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCCGAAGCA 407

Qy      606 GAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGCGCAAC 665
            ||||  |||||||||||||||||| |||||||||||||| || || ||||||||||||||
Db      406 GAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGGCGCAAC 347

Qy      666 AGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGGCCAAC 725
            ||||||||||||||||||||||| ||||| |||||| ||||||||||||||||||||||
Db      346 AGGCCCAAGAGCGAGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAGGCCAAT 287

Qy      726 GACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTC 785
            || |||||||||| |||||||||||||| ||||| ||||||||||||||||||||||||
Db      286 GATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGCGGTGTC 227

Qy      786 CTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCAC 845
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      226 CTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCAC 167

Qy      846 AACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCC 905
            |||||||| |||||||||||||||||||||||||||||||||||||||||| |||||
Db      166 AACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTATGGCCCG 107

Qy      906 GAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCGACTGG 965
            |||||||||||||||| || |||||||| ||||| ||| ||||||||| ||| ||||||
Db      106 GAGCCCCCTGTGCCCTG-GGAGAGCTGGGGAGCCCCGGA-GGGGCTCCAACGGGGACTGG 49

Qy      966 GGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCC 1010
            |||||     ||||||||| |||||||| || | |  | || |
Db       48 GGCTC--ATCACTCCCCAGTCTCCCAAGCGGGAACCTAGCCCCAC 6
```

RESULT 4
CA945402
LOCUS CA945402 687 bp mRNA linear EST 30-DEC-2002
DEFINITION UI-M-FD0-cdh-l-12-0-UI.r1 NIH_BMAP_FD0 Mus musculus cDNA clone IMAGE: 6828925 5', mRNA sequence.
ACCESSION CA945402
VERSION CA945402.1 GI:27433882
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 687)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, Univeristy of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone Distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
Seq primer: pYX-5.
FEATURES Location/Qualifiers
source 1. .687
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6828925"
/tissue_type="whole brain"
/dev_stage="embryo 12.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FD0"
/note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNa was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail , is TGAGAGAGCC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Instututes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT 117 a 228 c 250 g 91 t 1 others
ORIGIN

Query Match 9.1%; Score 133.4; DB 14; Length 687;
Best Local Similarity 83.1%; Pred. No. 5.1e-22;
Matches 152; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy      705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
            | |||  |    |||||||||| |||| |||||||||||||||||||||| ||| |||||
Db      422 ACCCGCAGGCTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCG 481

Qy      765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
            ||||| ||||||||||  || |||||||||||||| || || ||||| || || ||||||
Db      482 CTGGACGCGCTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATC 541

Qy      825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCG 884
            ||||| |||||||||||||||| ||||| |||||| || ||  |||| |||||| | |||
Db      542 GAGACGCTGCGCTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCG 601

Qy      885 GAC 887
            |||
Db      602 GAC 604
```

RESULT 5
AL540071
LOCUS AL540071 1001 bp mRNA linear EST 31-MAY-2003
DEFINITION AL540071 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF035YD19 5-PRIME, mRNA sequence.
ACCESSION AL540071
VERSION AL540071.2 GI:31264632
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1001)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

```
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12869886.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            2626.r, Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF035CB10QP1.
FEATURES             Location/Qualifiers
     source          1. .1001
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF035YD19"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
BASE COUNT      190 a    322 c    342 g    121 t     26 others
ORIGIN

  Query Match             9.1%;  Score 133;  DB 9;  Length 1001;
  Best Local Similarity   79.9%;  Pred. No. 7.5e-22;
  Matches  151;  Conservative    4;  Mismatches   34;  Indels    0;  Gaps    0;

Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
              | |||  |   ||||||||||| |||| ||||| |||||||||||||||||| ||| | ||
Db        706 ACCCGTAGACTGAAGGCCAACAACCGCGAGCGAAACCGCATGCACAACCTCAACGCGGCA 765

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
              ||||| |||||||||||  || || ||||| ||||| || |||||||| || || ||||||
Db        766 CTGGACGCGCTGCGCGAGGTGCTCCCCACGTTCCCCGAGGACGCCAAGCTCACCAAGATC 825

Qy        825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCG 884
              |||||:|||||||||||||||||||||||| |||| ||| ||  |||| |||||| | |||
Db        826 GAGACMCTGCGCTTCGCCCACAACTACATCTGGGGACTCACSGAGACCCTGCGCCTGGCG 885

Qy        885 GACCACAGC 893
              |: :|: ||
Db        886 GRTMAYTGC 894


RESULT 6
BJ010277
LOCUS       BJ010277                632 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ010277 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA142H01 5',
            mRNA sequence.
ACCESSION   BJ010277
VERSION     BJ010277.1  GI:17364159
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA142H01"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
BASE COUNT      120 a    231 c    190 g     90 t      1 others
ORIGIN

  Query Match             8.9%;  Score 130.2;  DB 12;  Length 632;
  Best Local Similarity   72.7%;  Pred. No. 3e-21;
  Matches  168;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy        658 GGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGA 717
              | ||   | || |      |||||  ||   ||      |  |  |   | ||||||   ||
Db        248 GCCGAGGCCGGGCGCGCTGCGAGGCGGTCGTGCAGGTGGTGAAAAAGAACCGGCGGCTGA 307

Qy        718 AGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGC 777
              |||| |||||||||| || |||||||||| ||||| |||  |||||||||| |||||
Db        308 AGGCTAACGACCGGGAACGGAACCGCATGCATAACCTGAACGACGCGCTGGATGAGCTGC 367

Qy        778 GCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCT 837
              |||| |||||||| |||| ||||| ||  |||| || || |||||||||| |||||||
Db        368 GCGGCGTCCTGCCGGCCTTTCCGGACGAAACCAAGCTGACCAAGATCGAGACTCTGCGCT 427

Qy        838 TCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACC 888
              | || |||||||||| ||||| ||| |  ||||  | ||||| || ||||
Db        428 TTGCGCACAACTACATCTGGGCTCTGTCCGAGACCATCCGCATCGCAGACC 478


RESULT 7
BG808248
LOCUS       BG808248                600 bp    mRNA    linear   EST 20-DEC-2001
DEFINITION  2083-52 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
            mRNA sequence.
ACCESSION   BG808248
VERSION     BG808248.1  GI:17955225
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
            White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
  TITLE     Gene expression in the developing mouse retina by EST sequencing
            and microarray analysis
  JOURNAL   Nucleic Acids Res. 29 (24), 4983-4993 (2001)
  MEDLINE   21671825
   PUBMED   11812828
COMMENT     Contact: Klein WH
            Department of Biochemistry and Molecular Biology
            University of Texas M.D. Anderson Cancer Center
            Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
            Tel: 713 792 3646
            Fax: 713 790 0329.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="neural retina"
                     /dev_stage="embryonic day 14.5 post-fertilization"
                     /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
BASE COUNT       94 a    238 c    161 g    107 t
ORIGIN
```

```
  Query Match             8.8%;  Score 128.6;  DB 12;  Length 600;
  Best Local Similarity   79.6%;  Pred. No. 7.3e-21;
  Matches  152;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy        729 CGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTG 788
              || |||||||||||||||||||||| ||| |||||||||| ||||||||||  || |||
Db          1 CGCGAGCGCAACCGCATGCACAACCTAAACGCCGCGCTGGACGCGCTGCGCGAGGTGCTG 60

Qy        789 CCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCGCCCACAAC 848
              |||||||||| || || ||||| || || |||||||||| ||||||||||||||||||
Db         61 CCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATCGAGACGCTGCGCTTCGCCCACAAT 120

Qy        849 TACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAG 908
              ||||| ||||| || ||  |||| |||||| | |||||||||| ||  |  || |   |
Db        121 TACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCGGACCACTGCGCCGGCGCCGGTGGC 180

Qy        909 CCCCCTGTGCC 919
              | ||  | | |
Db        181 CTCCAGGGGGC 191


RESULT 8
BU612495
LOCUS       BU612495                730 bp    mRNA    linear   EST 20-FEB-2003
DEFINITION  UI-M-FR0-cbc-k-21-0-UI.r1 NIH_BMAP_FR0 Mus musculus cDNA clone
            UI-M-FR0-cbc-k-21-0-UI 5', mRNA sequence.
ACCESSION   BU612495
VERSION     BU612495.1  GI:23278710
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 730)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .730
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="UI-M-FR0-cbc-k-21-0-UI"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FR0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      131 a    244 c    264 g     91 t
ORIGIN

  Query Match             8.8%;  Score 128.6;  DB 13;  Length 730;
  Best Local Similarity   81.4%;  Pred. No. 7.9e-21;
  Matches  149;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
              | |||  |    ||||||||| |||| ||||||||||||||||||||| ||| |||||
Db        543 ACCCGCAGGCTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCG 602

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
              ||||| ||||||||||  || ||||||||||||| || || ||||| || || ||||||
Db        603 CTGGACGCGCTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATC 662

Qy        825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCG 884
              ||||| ||||||||||||   | ||||| ||||| || ||  |||| |||||| | |||
Db        663 GAGACGCTGCGCTTCGCCCCACATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCG 722

Qy        885 GAC 887
              |||
Db        723 GAC 725


RESULT 9
BU054481
LOCUS       BU054481                823 bp    mRNA    linear   EST 26-AUG-2002
DEFINITION  UI-M-FD0-bzj-i-24-0-UI.r1 NIH_BMAP_FD0 Mus musculus cDNA clone
            IMAGE:6404447 5', mRNA sequence.
ACCESSION   BU054481
VERSION     BU054481.1  GI:22494558
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 823)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .823
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6404447"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FD0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
```

cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail , is TGAGAGAGCC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Instututes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 155 a 273 c 289 g 104 t 2 others
ORIGIN

Query Match 8.7%; Score 127.4; DB 13; Length 823;
Best Local Similarity 81.6%; Pred. No. 1.6e-20;
Matches 146; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
              | |||  |    ||||||||| |||| ||||||||||||||||||||||| ||| |||||
Db        645 ACCCGCAGGCTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCG 704

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
              ||||| ||||||||||  || ||||||||||||||| || || ||||  || || | ||||
Db        705 CTGGACGCGCTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCANGCTCACGANGATC 764

Qy        825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGC 883
              ||||| |||||||||||||||| ||||| ||||| || ||  |||| |||||| | ||
Db        765 GAGACGCTGCGCTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGC 823
```

RESULT 10
BX419330
LOCUS BX419330 947 bp mRNA linear EST 15-MAY-2003
DEFINITION BX419330 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF015YD06 5-PRIME, mRNA sequence.
ACCESSION BX419330
VERSION BX419330.1 GI:30765873
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 947)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 2626.r
Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DF015DB03QP1.
FEATURES Location/Qualifiers
source 1. .947
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF015YD06"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
BASE COUNT 73 a 393 c 315 g 130 t 36 others
ORIGIN

Query Match 8.3%; Score 121.8; DB 13; Length 947;
Best Local Similarity 72.7%; Pred. No. 4.1e-19;
Matches 152; Conservative 12; Mismatches 44; Indels 1; Gaps 1;

```
Qy        704 AAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGC 763
              :| |||  |   |||||||||: |:|| ||||| ||||||||||||||| ||| | ||
Db        717 RACCCGTAGACTGAAGGCCAAMAAMCGCGAGCGAAACCGCATGCACAACCTCAACGCGGC 776

Qy        764 GCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGAT 823
              ||||| |||||||||| :|| || ||||| ||||| || ||||||||:|| || || ||
Db        777 ACTGGACGCGCTGCGCGAKGTGCTCCCCACGTTCCCCGAGGACGCCAARCTCACCAAAAT 836

Qy        824 CGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGC 883
              :||||:|||||||||| |||:|| :| |: |||||||| ||  |:|| |||||| | ||
Db        837 MGAGAMCCTGCGCTTCG-CCAMAAAWAAAWMTGGGCACTCACCGARACCCTGCGCCTGGC 895

Qy        884 GGACCACAGCTTCTACGGCCCCGAGCCCC 912
              ||| ||| ||     :|||  || |  ||
Db        896 GGATCACTGCGGGGGSGGCGGCGGGGGCC 924
```

RESULT 11
BU924937
LOCUS BU924937 600 bp mRNA linear EST 30-OCT-2002
DEFINITION 7103-91 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA, mRNA sequence.
ACCESSION BU924937
VERSION BU924937.1 GI:24428820
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 600)
AUTHORS Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W., White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
TITLE Gene expression in the developing mouse retina by EST sequencing and microarray analysis
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)
MEDLINE 21671825
PUBMED 11812828
COMMENT Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329.
FEATURES Location/Qualifiers
source 1. .600
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
BASE COUNT 99 a 194 c 222 g 82 t 3 others
ORIGIN

Query Match 8.2%; Score 120.4; DB 13; Length 600;
Best Local Similarity 75.1%; Pred. No. 7.5e-19;
Matches 163; Conservative 0; Mismatches 52; Indels 2; Gaps 1;

```
Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
              | |||  |    ||||||||| |||| ||||||||||||||||||||||| ||| |||||
Db        336 ACCCGCAGGCTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCG 395

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
              ||||| ||||||||||  || ||||||||||||||| || || ||||| || || ||||||
Db        396 CTGGACGCGCTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATC 455

Qy        825 GAGACCCT--GCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAG 882
              ||||| ||      |||||||||| ||||| ||||| || ||  |||| |||||| | |
Db        456 GAGACGCTTGCGCTTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGG 515

Qy        883 CGGACCACAGCTTCTACGGCCCCGAGCCCCCTGTGCC 919
              |||||||| ||  | || |  |  |  |  | | |
Db        516 CGGACCACTGCGCCGGCGCCGGTGGCCTNCAGGGGGC 552
```

```
RESULT 12
CA979119
LOCUS       CA979119                 932 bp    mRNA    linear   EST 06-JAN-2003
DEFINITION  AGENCOURT_11295215 NIH_MGC_164 Mus musculus cDNA clone
            IMAGE:30146192 5', mRNA sequence.
ACCESSION   CA979119
VERSION     CA979119.1  GI:27511773
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe and Dr. Mina
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM0061  row: h  column: 09
            High quality sequence start: 16
            High quality sequence stop: 640.
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30146192"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_164"
                     /note="Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2:
                     NotI; Non-normalized full-length enriched library from
                     pooled mouse embryonic limb, maxilla and mandible, day
                     10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
                     Cloned directionally, priming method: Oligo-dT.  cDNA
                     enrichment: >1k bp, Average insert size 1.8k bp. Priming
                     sequence: 5'GACTAGTTCTAGATCGCGAGCGGCCGCCC(T) 3'. Tissue
                     contributed by, David Rowe. Library constructed by ResGen,
                     Invitrogen Corp."
BASE COUNT      153 a    349 c    276 g    153 t      1 others
ORIGIN

  Query Match             8.1%;  Score 118.6;  DB 14;  Length 932;
  Best Local Similarity   66.1%;  Pred. No. 2.5e-18;
  Matches  187;  Conservative    0;  Mismatches   94;  Indels    2;  Gaps    1;

Qy        635 GAAGCTCCGTGCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAA 694
              | ||   ||   | ||||  |||| |||    ||      |  ||||      |||  |
Db        464 GCAGGAACGGCGGAGGCGGCGAGGTCGCGCTCGGGTGCGGTCCGAGGCTCTGCTGCACTC 523

Qy        695 GCAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCT 754
              | |||  | || || |||   || |||||||| || ||||||||||||||||||||||
Db        524 CCTGCGGAGGAGTCGTCGCGTCAAAGCCAACGATCGCGAGCGCAACCGCATGCACAACCT 583

Qy        755 TAACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACT 814
              ||| | |||||||| ||  ||||| | || |||||| | ||||| || ||| |||| ||
Db        584 CAACGCTGCGCTGGACGCCTTGCGCAGCGTGCTGCCCTCGTTCCCCGACGACACCAAGCT 643

Qy        815 TACAAAGATCGAGACCCTGCGCTTCGCCCACAA--CTACATTTGGGCACTGACTCAGACG 872
              || ||||| ||||| |||||||||||| ||||        | |||| ||| || ||||
Db        644 CACCAAGATTGAGACGCTGCGCTTCGCCTACAAACTACATCTGGGGCCCTGGCTGAGACA 703

Qy        873 CTGCGCATAGCGGACCACAGCTTCTACGGCCCCGAGCCCCCTG 915
              |||||| | || || ||  |  ||  |||   |    |||| |
Db        704 CTGCGCCTGGCAGATCAAGGGCTCCCCGGGGGCAGTGCCCCGG 746
```

```
RESULT 13
CNS03VJ9/c
LOCUS       CNS03VJ9                 1037 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            062M23 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL262494
VERSION     AL262494.1  GI:7984120
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1037)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1037
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="062M23"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG062AG12LP1~end : T7"
BASE COUNT      205 a    254 c    267 g    308 t      3 others
ORIGIN

  Query Match             7.9%;  Score 115.6;  DB 29;  Length 1037;
  Best Local Similarity   74.7%;  Pred. No. 1.4e-17;
  Matches  145;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        696 CAGCGACGAAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTT 755
              |||||  |  | ||  | | ||| || |||||| |||||||| |||| ||||||||||
Db        417 CAGCGGGGCCGACGCAGGATGAAAGCGAACGACAGGGAGCGCCACCGGATGCACAACCTG 358

Qy        756 AACTCCGCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTT 815
              |||||||| ||||| |||||| | ||  |||||||      | || || ||| | || ||
Db        357 AACTCCGCCCTGGACGCGCTGAGGGGCATCCTGCCGGTGCTACCCGAGGACACAAAGCTG 298

Qy        816 ACAAAGATCGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTG 875
              || || || || || ||||| || |||||||||||||| |||||  ||||  |||| |||
Db        297 ACCAAAATTGAAACTCTGCGTTTTGCCCACAACTACATCTGGGCTTTGACCGAGACTCTG 238

Qy        876 CGCATAGCGGACCA 889
```

```
              ||||| || || ||
Db        237 CGCATGGCCGATCA 224


RESULT 14
AW147434
LOCUS       AW147434                401 bp    mRNA    linear   EST 19-FEB-2003
DEFINITION  da02h12.y1 Xenopus laevis oocyte Xenopus laevis cDNA clone
            XENOPUS_SOURCE_ID:xlnoc001h24 5' similar to gb:gb|U67779.1|XLU67779
            Xenopus laevis neurogenin related 1b (XENOPUS);, mRNA sequence.
ACCESSION   AW147434
VERSION     AW147434.1  GI:6195330
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Bruce Blumberg
             Library normalized by Jihwan Song
             DNA Sequencing by: Washington University Genome Sequencing Center
            Seq primer: -40RP from Gibco
            High quality sequence stop: 264.
FEATURES             Location/Qualifiers
     source          1. .401
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="XENOPUS_SOURCE_ID:xlnoc001h24"
                     /tissue_type="oocyte (stages 5 and 6)"
                     /lab_host="Top-10 F'"
                     /clone_lib="Xenopus laevis oocyte"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from 2ug of poly A+ RNA.
                     EcoRI-XhoI cut cDNA was then ligated into UniZap-XR
                     (Stratagene) with EcoRI at the 5' end and XhoI at the 3'
                     end. SS-library phagemids were prepared by mass excision
                     from the original library and normalized by hybridization
                     to biotinylated driver (prepared from the same library by
                     PCR) to Cot-omega of 11. After removal of hybrids and
                     excess driver by streptavidin sepharose chromatography,
                     the ss-phagemids were made double stranded and
                     electroporated into Top-10 F'. Original library
                     contruction by Bruce Blumberg (Blumberg et al., 1991
                     Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,
                     2923-2935). Normalized by Jihwan Song (Song, Cho and
                     Blumberg, unpublished). Note: This is a Xenopus Gene
                     Collection (XGC) library."
BASE COUNT       87 a    140 c     84 g     90 t
ORIGIN

  Query Match             7.6%;  Score 110.8;  DB 9;  Length 401;
  Best Local Similarity   74.7%;  Pred. No. 1.4e-16;
  Matches  139;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy        704 AAGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGC 763
              || |||||||   || |||||  |||| ||| | || |||||||| |||| ||||  ||
Db          4 AACCCGGCGCGTTAAAGCCAATAACCGCGAGAGGAATCGCATGCACCACCTGAACTATGC 63

Qy        764 GCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGAT 823
              ||| ||| | ||| | |  || || ||  | || || || |||||||||| || |||||
Db         64 GCTCGATTCTCTGAGGGAGGTTCTACCGTCATTACCCGAAGACGCCAAACTCACCAAGAT 123

Qy        824 CGAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGC 883
              |||||| ||||||| |||||||||||||| ||||| || |   | ||  ||||| | ||
Db        124 AGAGACCTTGCGCTTTGCCCACAACTACATCTGGGCTCTTAGCGAAACTTTGCGCCTGGC 183

Qy        884 GGACCA 889
              ||||||
Db        184 CGACCA 189


RESULT 15
CD282259
LOCUS       CD282259                603 bp    mRNA    linear   EST 23-MAY-2003
DEFINITION  G39271.16 NCI_CGAP_ZEmb2 Danio rerio cDNA clone IMAGE:6521499 5',
            mRNA sequence.
ACCESSION   CD282259
VERSION     CD282259.1  GI:31060035
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Amundsen,C., Cachuela,N., Chen,F., Cheung,L.M., Chong,A., Murray,L.
            , Oliva,J., Park,C., Reyes,J., Yungen,J. and Swimmer,C.
  TITLE     Expressed sequence tags from NCI_CGAP_ZEmb2, a Danio rerio
            embryonic library
  JOURNAL   Unpublished
COMMENT     Contact: Chen F.
            Exelixis, Inc.
            170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA
            Tel: 650 837 7000
            Fax: 650 837 8300
            Email: fchen@exelixis.com
            DNA Sequencing by: Exelixis, Inc. Clone distribution information
            can be found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: 14108  row: O  column: 3
            High quality sequence stop: 603.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:6521499"
                     /tissue_type="embryo"
                     /lab_host="DH10B (T1-resistant)"
                     /clone_lib="NCI_CGAP_ZEmb2"
                     /note="Vector: pCMV-SPORT6.ccdb; Site_1: EcoRV; Site_2:
                     NotI; Cloned unidirectionally. Primer: Oligo dT. Average
                     insert size 2 kb. Constructed by J. Wang (Research
                     Genetics, Invitrogen Corp) from tissue donated by L. Zon
                     (Harvard University). Note: this is a NCI_CGAP Library."
BASE COUNT      159 a    185 c    155 g    104 t
ORIGIN

  Query Match             7.5%;  Score 110.2;  DB 14;  Length 603;
  Best Local Similarity   74.3%;  Pred. No. 2.4e-16;
  Matches  139;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        705 AGCCGGCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCG 764
              | |||  |  ||||||||||||||| ||| | ||| | |||||||||||||  |||
Db        325 AACCGCAGGCTGAAGGCCAACGACCGCGAGAGGAACAGGATGCACAACCTTAACGACGCA 384

Qy        765 CTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATC 824
              |||||||  || |  | ||||||||  | || || || ||| | || || || || ||
Db        385 TTGGATGCTTTGAGAAGCGTCCTGCCTGCGTTTCCTGACGACACAAAGCTGACCAAAATT 444

Qy        825 GAGACCCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCG 884
```

```
                ||||| |||||||||||| |||||||||| ||||||||  |  ||||  | || || ||
Db        445 GAGACTCTGCGCTTCGCTCACAACTACATCTGGGCACTTTCGGAGACCATCCGGATCGCA 504

Qy        885 GACCACA 891
                ||||| |
Db        505 GACCAGA 511
```

Search completed: January 26, 2004, 22:05:05
Job time : 3312 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 18:54:03 ; Search time 5306 Seconds
(without alignments)
11256.701 Million cell updates/sec

Title: US-09-595-947E-1
Perfect score: 1460
Sequence: 1 gcaggtagcgagaggagcag.....agagtgacctaatccagtgt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1409	96.5	1491	6	A91167	A91167 Sequence 1
2		1409	96.5	1491	6	BD023626	BD023626 Polypepti
3		1409	96.5	1491	10	RNRELAXT	Y10619 R.norvegicu
4		1060	72.6	258815	2	AC127817	AC127817 Rattus no
5		92	6.3	861	6	AX698801	AX698801 Sequence
6		92	6.3	861	10	MMU76208	U76208 Mus musculu
7		92	6.3	1861	10	AF364300	AF364300 Mus muscu
8		92	6.3	5567	10	MMMATH4B	Y09167 M.musculus
9	c	92	6.3	138070	2	AC109783	AC109783 Mus muscu
10		92	6.3	215050	2	AC127417	AC127417 Mus muscu
11		35	2.4	2370	9	BC036847	BC036847 Homo sapi
12		35	2.4	6123	9	AF303002	AF303002 Homo sapi
13	c	35	2.4	91531	2	AC079846_3	Continuation (4 of
14	c	35	2.4	179697	9	AC023886	AC023886 Homo sapi
15		32	2.2	1330	9	HSA133776	AJ133776 Homo sapi
16		32	2.2	5340	9	AF234829	AF234829 Homo sapi
17	c	32	2.2	165110	9	AL450311	AL450311 Human DNA
18	c	32	2.2	173341	2	AC021954	AC021954 Homo sapi
19	c	26	1.8	26	6	AX548297	AX548297 Sequence
20	c	26	1.8	170896	2	AC011010	AC011010 Homo sapi
21		25	1.7	25	6	A91170	A91170 Sequence 4
22		25	1.7	25	6	BD023629	BD023629 Polypepti
23		23	1.6	735	10	MMU67776	U67776 Mus musculu
24		23	1.6	738	6	AR308548	AR308548 Sequence
25		23	1.6	770	5	AF123884	AF123884 Gallus ga
26		23	1.6	790	5	GGA012659	AJ012659 Gallus ga
27		23	1.6	932	10	MMMATH4C	Y09166 M.musculus
28		23	1.6	1315	10	MMU63841	U63841 Mus musculu
29		23	1.6	1333	6	AR023715	AR023715 Sequence
30		23	1.6	1333	6	AR225848	AR225848 Sequence
31		23	1.6	1341	5	AF109014	AF109014 Gallus ga
32		23	1.6	1385	10	MMU76207	U76207 Mus musculu
33		23	1.6	1412	10	MMMATH4A	Y07621 M.musculus
34		23	1.6	1880	5	AF303000	AF303000 Gallus ga
35		23	1.6	10393	10	AF303001	AF303001 Mus muscu
36		23	1.6	71538	2	AC118243	AC118243 Mus muscu
37		23	1.6	123855	2	AC102600	AC102600 Mus muscu
38		23	1.6	149268	2	AC124395	AC124395 Mus muscu
39		23	1.6	262798	2	AC111702	AC111702 Rattus no
40	c	22	1.5	215591	2	AC102159	AC102159 Mus muscu
41		21	1.4	21	6	AX553597	AX553597 Sequence
42		21	1.4	21	6	AX553645	AX553645 Sequence
43		21	1.4	310	6	I84656	I84656 Sequence 12
44		21	1.4	1058	9	BC052235	BC052235 Homo sapi
45		21	1.4	1164	5	GGNEURODL	Y09596 G.gallus mR

ALIGNMENTS

RESULT 1
A91167
LOCUS A91167 1491 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent WO9827206.
ACCESSION A91167
VERSION A91167.1 GI:6740202
KEYWORDS .
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1491)
AUTHORS Icard-Liepkalns,C., Mallet,J. and Corresponding,N.A.
JOURNAL Patent: WO 9827206-A 1 25-JUN-1998;

ICARD LIEPKALNS CHRISTINE (FR); MALLET JACQUES (FR)

```
FEATURES             Location/Qualifiers
     source          1. .1491
                     /organism="Rattus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10118"
     CDS             459. .1103
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAB69414.1"
                     /db_xref="GI:6740203"
                     /translation="MAPHPLDAPTIQVSQETQQPFPGASDHEVLSSNSTPPSPTLVPR
                     DCSEAEAGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSSGDWGSIYSPVSQAGSLSPTASLEEFPGLQVPSSPSCLLPGTLVFSDFL"
BASE COUNT      307 a    487 c    413 g    284 t
ORIGIN
```

Query Match 96.5%; Score 1409; DB 6; Length 1491;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60

Qy         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120

Qy        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180

Qy        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240

Qy        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300

Qy        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360

Qy        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420

Qy        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480

Qy        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540

Qy        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600

Qy        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660

Qy        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720

Qy        721 CCAACGACCGGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780

Qy        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840

Qy        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900

Qy        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960

Qy        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020

Qy       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080

Qy       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140

Qy       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200

Qy       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260

Qy       1261 GCTGTGCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       1261 GCTGTGCGTGCACAAAGGACATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320

Qy       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380

Qy       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440

Qy       1441 AGAGTGACCTAATCCAGTGT 1460
              ||||||||||||||||||||
Db       1441 AGAGTGACCTAATCCAGTGT 1460
```

RESULT 2
BD023626

```
LOCUS       BD023626               1491 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  Polypeptide belonging to the family of basic helix-loop-helix
            (bHLH) family and nucleic acid sequence corresponding thereto.
ACCESSION   BD023626
VERSION     BD023626.1  GI:22564849
KEYWORDS    JP 2001510464-A/1.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1491)
  AUTHORS   Liepkalns,C.I., Mallet,J. and Ravassard,P.
  TITLE     Polypeptide belonging to the family of basic helix-loop-helix
            (bHLH) family and nucleic acid sequence corresponding thereto
  JOURNAL   Patent: JP 2001510464-A 1 31-JUL-2001;
            RHONE POULENC RORER SA
COMMENT     OS   Rattus sp. (rat)
            PN   JP 2001510464-A/1
            PD   31-JUL-2001
            PF   19-DEC-1997 JP 1998527415
```

```
            PR   19-DEC-1996 FR     96/15651
            PI   CHRISTINE ICARD LIEPKALNS,JACQUES MALLET,PHILIPPE RAVASSARD PC
              C07K14/47,A61K31/711,A61K35/76,A61K38/00,A61K48/00,A61P25/00, PC
             C12N15/09,
            PC   C12N15/00,A61K37/02
            CC   Strandedness: Single;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES             Location/Qualifiers
     source          1. .1491
                     /organism="Rattus sp."
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10118"
BASE COUNT      307 a    487 c    413 g    284 t
ORIGIN

  Query Match             96.5%;  Score 1409;  DB 6;  Length 1491;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1459;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60

Qy         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120

Qy        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180

Qy        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240

Qy        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300

Qy        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360

Qy        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420

Qy        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480

Qy        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540

Qy        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600

Qy        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660

Qy        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720

Qy        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780

Qy        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840

Qy        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900

Qy        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960

Qy        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020

Qy       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080

Qy       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140

Qy       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200

Qy       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260

Qy       1261 GCTGTGCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTGTGCGTGCACAAAGGACATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320

Qy       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380

Qy       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440

Qy       1441 AGAGTGACCTAATCCAGTGT 1460
              ||||||||||||||||||||
Db       1441 AGAGTGACCTAATCCAGTGT 1460


RESULT 3
RNRELAXT
LOCUS       RNRELAXT                1491 bp    mRNA    linear   ROD 06-MAY-1997
DEFINITION  R.norvegicus mRNA for transcriptional regulator, Relax.
ACCESSION   Y10619
VERSION     Y10619.1  GI:2072737
KEYWORDS    Relax; transcriptional regulator.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Ravassard,P., Chatail,F., Mallet,J. and Icard-Liepkalns,C.
  TITLE     Relax, a novel rat bHLH transcriptional regulator transiently
            expressed in the ventricular proliferating zone of the developing
            central nervous system
  JOURNAL   J. Neurosci. Res. 48 (2), 146-158 (1997)
  MEDLINE   97276390
   PUBMED   9130143
REFERENCE   2  (bases 1 to 1491)
  AUTHORS   Ravassard,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,
            Hopital de la Pitie Salpetriere, 83 Bd. de l'Hopital, F-75013
```

```
          Paris, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1491
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /clone_lib="RT-PCR"
     CDS             459. .1103
                     /function="transcriptional regulator"
                     /codon_start=1
                     /product="Relax"
                     /protein_id="CAA71630.1"
                     /db_xref="GI:2072738"
                     /db_xref="SPTREMBL:O08718"
                     /translation="MAPHPLDAPTIQVSQETQQPFPGASDHEVLSSNSTPPSPTLVPR
                     DCSEAEAGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSSGDWGSIYSPVSQAGSLSPTASLEEFPGLQVPSSPSCLLPGTLVFSDFL"
BASE COUNT      307 a    487 c    413 g    284 t
ORIGIN

  Query Match             96.5%;  Score 1409;  DB 10;  Length 1491;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1459;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60

Qy         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120

Qy        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180

Qy        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240

Qy        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300

Qy        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360

Qy        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420

Qy        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480

Qy        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540

Qy        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600

Qy        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660

Qy        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720

Qy        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780

Qy        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840

Qy        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900

Qy        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960

Qy        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020

Qy       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080

Qy       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140

Qy       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200

Qy       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260

Qy       1261 GCTGTGCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTGTGCGTGCACAAAGGACATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320

Qy       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380

Qy       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440

Qy       1441 AGAGTGACCTAATCCAGTGT 1460
              ||||||||||||||||||||
Db       1441 AGAGTGACCTAATCCAGTGT 1460


RESULT 4
AC127817
LOCUS       AC127817               258815 bp    DNA     linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-259G16, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC127817
VERSION     AC127817.3  GI:25077905
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 258815)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
```

Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 258815)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 258815)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Nov 19, 2002 this sequence version replaced gi:23912578.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GZXS
Center clone name: CH230-259G16
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 224747 bases at least Q40
Consensus quality: 227981 bases at least Q30
Consensus quality: 229752 bases at least Q20
Estimated insert size: 228243; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 255979: contig of 255979 bp in length
* 255980 256079: gap of unknown length
* 256080 257349: contig of 1270 bp in length
* 257350 257449: gap of unknown length
* 257450 258815: contig of 1366 bp in length.

FEATURES Location/Qualifiers

source 1. .258815
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-259G16"

misc_feature 1. .1609
/note="wgs_end_extension
clone_end:Sp6"

misc_feature complement(2865. .3659)
/note="clone_boundary
clone_end:Sp6
site:
end_sequence:BZ169379"

misc_feature 224098. .225157
/note="clone_boundary
clone_end:T7
site:
end_sequence:BZ169378"

BASE COUNT 59351 a 56888 c 58131 g 56821 t 27624 others

ORIGIN

Query Match 72.6%; Score 1060; DB 2; Length 258815;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy        350 CCTCCAAAGGGTAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCT 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247751 CCTCCAAAGGGTAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCT 247810

Qy        410 CTAAGTCAGAGACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247811 CTAAGTCAGAGACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCA 247870

Qy        470 TCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247871 TCCCTTGGATGCGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGC 247930

Qy        530 CTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     247931 CTCGGACCACGAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAG 247990

Qy        590 GGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      247991 GGACTGCTCCGAAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCG 248050

Qy         650 GCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCG 709
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248051 GCGCGGAGGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCG 248110

Qy         710 GCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGA 769
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248111 GCGCAAGAAGGCCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGA 248170

Qy         770 TGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGAC 829
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248171 TGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGAC 248230

Qy         830 CCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCA 889
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248231 CCTGCGCTTCGCCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCA 248290

Qy         890 CAGCTTCTACGGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGG 949
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248291 CAGCTTCTACGGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGG 248350

Qy         950 CTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCC 1009
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248351 CTCCAGCGGCGACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCC 248410

Qy        1010 CACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCT 1069
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248411 CACAGCCTCATTGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCT 248470

Qy        1070 CCCGGGCACCCTGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTG 1129
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248471 CCCGGGCACCCTGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTG 248530

Qy        1130 GGCGCTGGCAGAAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTC 1189
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248531 GGCGCTGGCAGAAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTC 248590

Qy        1190 GAGCATCTCGCCCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCAC 1249
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248591 GAGCATCTCGCCCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCAC 248650

Qy        1250 AGTTCCTTGCTGCTGTGCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCC 1309
               ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      248651 AGTTCCTTGCTGCTGTGCGTGCACAAAGGACATTGCAGGCTGATCTCCTCTTAACCCTCC 248710

Qy        1310 TCAGTGTGGCCACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTT 1369
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248711 TCAGTGTGGCCACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTT 248770

Qy        1370 GGGAGACTCCCATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCA 1429
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      248771 GGGAGACTCCCATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCA 248830

Qy        1430 CCGCTTTCTCCAGAGTGACCTAATCCAGTGT 1460
               |||||||||||||||||||||||||||||||
Db      248831 CCGCTTTCTCCAGAGTGACCTAATCCAGTGT 248861


RESULT 5
AX698801
LOCUS       AX698801                861 bp    DNA     linear   PAT 02-APR-2003
DEFINITION  Sequence 7 from Patent WO02086107.
ACCESSION   AX698801
VERSION     AX698801.1  GI:29499589
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Wobus,A.M., St-Onge,L., Blyszczuk,P. and Hoffmann,U.
  TITLE     A method for differentiating stem cells into insulin-producing
            cells
  JOURNAL   Patent: WO 02086107-A 7 31-OCT-2002;
            DeveloGen Aktiengesellschaft fuer Entwicklungsbiologische Forschung
            (DE) ; INSTITUT FUER PFLANZENGENETIK UND KULTURPFLANZENFORSCHUNG
            (DE)
FEATURES             Location/Qualifiers
     source          1. .861
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     CDS             160. .804
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD87442.1"
                     /db_xref="GI:29499590"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT      182 a    274 c    250 g    155 t
ORIGIN

  Query Match             6.3%;  Score 92;  DB 6;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-42;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         463 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 522

Qy         822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
               ||||||||||||||||||||||||||||||||
Db         523 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 554


RESULT 6
MMU76208
LOCUS       MMU76208                861 bp    DNA     linear   ROD 05-FEB-1997
DEFINITION  Mus musculus neurogenin 3 (ngn3) gene, complete cds.
ACCESSION   U76208
VERSION     U76208.1  GI:1815654
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 861)
  AUTHORS   Sommer,L., Ma,Q. and Anderson,D.J.
  TITLE     neurogenins, a novel family of atonal-related bHLH transcription
            factors, are putative mammalian neuronal determination genes that
            reveal progenitor cell heterogeneity in the developing CNS and PNS
  JOURNAL   Mol. Cell. Neurosci. 8 (4), 221-241 (1996)
  MEDLINE   97153565
   PUBMED   9000438
REFERENCE   2  (bases 1 to 861)
  AUTHORS   Sommer,L., Ma,Q. and Anderson,D.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-1996) Biology 216-76, California Institute of
            Technology, Howard Hughes Medical Institute, Pasadena, CA 91125,
            USA
COMMENT     On Feb 5, 1997 this sequence version replaced gi:1666911.
FEATURES             Location/Qualifiers
     source          1. .861
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene            160. .804
                     /gene="ngn3"
     CDS             160. .804
                     /gene="ngn3"
                     /function="bHLH transcription factor"
                     /note="NGN3"
                     /codon_start=1
```

```
                     /product="neurogenin 3"
                     /protein_id="AAC53029.1"
                     /db_xref="GI:1666912"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT       182 a    274 c    250 g    155 t
ORIGIN

  Query Match             6.3%;  Score 92;  DB 10;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-42;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 522

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||
Db        523 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 554


RESULT 7
AF364300
LOCUS       AF364300                1861 bp    DNA     linear   ROD 03-MAY-2001
DEFINITION  Mus musculus neurogenin 3 gene, complete cds.
ACCESSION   AF364300
VERSION     AF364300.1  GI:13937128
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1861)
  AUTHORS   Lee,J., Smith,S., Watada,H., Lin,J., Scheel,D., Wang,J., Mirmira,R.
            and German,M.
  TITLE     Regulation of the pancreatic pro-endocrine gene neurogenin3
  JOURNAL   Diabetes (2001) In press
REFERENCE   2  (bases 1 to 1861)
  AUTHORS   Schwitzgebel,V. and German,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2001) Hormone Research Institute, University of
            California San Francisco, 513 Parnassus Ave, HSW1090, San
            Francisco, CA 94145-0534, USA
FEATURES             Location/Qualifiers
     source          1. .1861
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     mRNA            <1093. .>1737
                     /product="neurogenin 3"
     CDS             1093. .1737
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="AAK50058.1"
                     /db_xref="GI:13937129"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT       397 a    560 c    537 g    367 t
ORIGIN

  Query Match             6.3%;  Score 92;  DB 10;  Length 1861;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-42;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1396 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 1455

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              ||||||||||||||||||||||||||||||||
Db       1456 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 1487


RESULT 8
MMMATH4B
LOCUS       MMMATH4B                5567 bp    DNA     linear   ROD 25-OCT-2000
DEFINITION  M.musculus MATH4B gene.
ACCESSION   Y09167
VERSION     Y09167.2  GI:11065737
KEYWORDS    bHLH; mammalian atonal homolog 4B; MATH4B; neurogenin 3.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Cau,E., Gradwohl,G., Fode,C. and Guillemot,F.
  TITLE     Mash1 activates a cascade of bHLH regulators in olfactory neuron
            progenitors
  JOURNAL   Development 124 (8), 1611-1621 (1997)
  MEDLINE   97261963
   PUBMED   9108377
REFERENCE   2
  AUTHORS   Jacquemin,P., Durviaux,S.M., Jensen,J., Godfraind,C., Gradwohl,G.,
            Guillemot,F., Madsen,O.D., Carmeliet,P., Dewerchin,M., Collen,D.,
            Rousseau,G.G. and Lemaigre,F.P.
  TITLE     Transcription factor hepatocyte nuclear factor 6 regulates
            pancreatic endocrine cell differentiation and controls expression
            of the proendocrine gene ngn3
  JOURNAL   Mol. Cell. Biol. 20 (12), 4445-4454 (2000)
  MEDLINE   20285449
   PUBMED   10825208
REFERENCE   3
  AUTHORS   Gradwohl,G.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC,
            CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg,
            F-67404 ILLKIRCH cedex, FRANCE
  REMARK    Revised by [4]
REFERENCE   4  (bases 1 to 5567)
  AUTHORS   Lemaigre,F.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic
            Research Unit, Louvain University Medical School, Avenue Hippocrate
            75, box 7529, Brussels 1200, BELGIUM
COMMENT     On Oct 31, 2000 this sequence version replaced gi:1666087.
FEATURES             Location/Qualifiers
     source          1. .5567
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="SV129"
                     /db_xref="taxon:10090"
                     /clone="lambda GEM12 clone 19C"
                     /cell_line="D3"
                     /cell_type="embryonic stem cells"
                     /clone_lib="Genomic library No.62, MboI partial digests
                     cloned in BamHI"
                     /dev_stage="embryos"
     misc_binding    958
                     /bound_moiety="HNF6"
     misc_binding    3692
                     /bound_moiety="HNF6"
     gene            4145. .5567
                     /gene="MATH4B"
     TATA_signal     4145. .4151
                     /gene="MATH4B"
                     /note="putative"
     CDS             4923. .5567
                     /gene="MATH4B"
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="CAA70366.1"
                     /db_xref="GI:1666088"
```

```
                     /db_xref="SWISS-PROT:P70661"
                     /translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSPTLIPR
                     DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKQRRSRRKKANDRERNRMHNLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSFYGPEPPVPCGELGS
                     PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"
BASE COUNT     1271 a   1549 c   1564 g   1183 t
ORIGIN

  Query Match             6.3%;  Score 92;  DB 10;  Length 5567;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-42;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5226 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 5285

Qy         822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
               |||||||||||||||||||||||||||||||
Db        5286 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 5317


RESULT 9
AC109783/c
LOCUS       AC109783              138070 bp    DNA     linear   HTG 07-FEB-2002
DEFINITION  Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
            unordered pieces.
ACCESSION   AC109783
VERSION     AC109783.1  GI:18581594
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 138070)
  AUTHORS   McCombie,W.R., Baker,J.P., Balija,V., Dedhia,N.N., de la
            Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
            Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,
            Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
            Zutavern,T.
  TITLE     Mouse Genomic Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 138070)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
COMMENT     -------------- Genome Center --------------
            Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
            Laboratory
            Center code: CSHL
            Web site: http://www.cshl.org/genseq
            Contact:  mccombie@cshl.org
            ------------------- Project Information
            Project name: RP23-121F10
            Clone name: RP23-121F10
            Insert size: 173000; agarose-fp
            Insert size: 141616; sum-of-contigs
            Quality coverage: 4.00 in Q20 bases; agarose-fp
            Quality coverage: 3.70 in Q20 bases; sum-of-contigs
            ----------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17785: contig of 17785 bp in length
            *    17786    17874: gap of unknown length
            *    17875    30356: contig of 12482 bp in length
            *    30357    30444: gap of unknown length
            *    30445    42306: contig of 11862 bp in length
            *    42307    42394: gap of unknown length
            *    42395    53598: contig of 11204 bp in length
            *    53599    53686: gap of unknown length
            *    53687    64355: contig of 10669 bp in length
            *    64356    64443: gap of unknown length
            *    64444    74016: contig of 9573 bp in length
            *    74017    74104: gap of unknown length
            *    74105    83366: contig of 9262 bp in length
            *    83367    83454: gap of unknown length
            *    83455    92355: contig of 8901 bp in length
            *    92356    92443: gap of unknown length
            *    92444   100821: contig of 8378 bp in length
            *   100822   100909: gap of unknown length
            *   100910   107529: contig of 6620 bp in length
            *   107530   107617: gap of unknown length
            *   107618   114066: contig of 6449 bp in length
            *   114067   114154: gap of unknown length
            *   114155   118873: contig of 4719 bp in length
            *   118874   118961: gap of unknown length
            *   118962   123619: contig of 4658 bp in length
            *   123620   123707: gap of unknown length
            *   123708   128240: contig of 4533 bp in length
            *   128241   128328: gap of unknown length
            *   128329   132682: contig of 4354 bp in length
            *   132683   132770: gap of unknown length
            *   132771   136341: contig of 3571 bp in length
            *   136342   136429: gap of unknown length
            *   136430   138070: contig of 1641 bp in length.
FEATURES             Location/Qualifiers
     source          1. .138070
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-121F10"
BASE COUNT    35731 a  33657 c  33303 g  33954 t   1425 others
ORIGIN

  Query Match             6.3%;  Score 92;  DB 2;  Length 138070;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-43;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      110917 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 110858

Qy         822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
               |||||||||||||||||||||||||||||||
Db      110857 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 110826


RESULT 10
AC127417
LOCUS       AC127417              215050 bp    DNA     linear   HTG 19-OCT-2002
DEFINITION  Mus musculus chromosome UNK clone RP23-459M2, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC127417
VERSION     AC127417.2  GI:24137619
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 215050)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 215050)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 215050)
```

```
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Oct 19, 2002 this sequence version replaced gi:21759524.

            ---------------- Genome Center ----------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            -------------- Project Information ------------------
            Center project name: M_BA0459M02
            -----------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 212279 bases at least Q40
            Consensus quality: 213137 bases at least Q30
            Consensus quality: 213590 bases at least Q20
            Insert size: 192000; agarose-fp
            Insert size: 217944; sum-of-contigs
            Quality coverage: 11.89 in Q20 bases; agarose-fp
            Quality coverage: 10.54 in Q20 bases; sum-of-contigs
            -------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     4227: contig of 4227 bp in length
            *    4228     4327: gap of unknown length
            *    4328    16589: contig of 12262 bp in length
            *   16590    16689: gap of unknown length
            *   16690    27318: contig of 10629 bp in length
            *   27319    27418: gap of unknown length
            *   27419    49797: contig of 22379 bp in length
            *   49798    49897: gap of unknown length
            *   49898    79378: contig of 29481 bp in length
            *   79379    79478: gap of unknown length
            *   79479   132595: contig of 53117 bp in length
            *  132596   132695: gap of unknown length
            *  132696   215050: contig of 82355 bp in length.
FEATURES             Location/Qualifiers
     source          1. .215050
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="UNK"
                     /clone="RP23-459M2"
     misc_feature    1. .4227
                     /note="assembly_name:Contig31"
     misc_feature    4328. .16589
                     /note="assembly_name:Contig32"
     misc_feature    16690. .27318
                     /note="assembly_name:Contig33"
     misc_feature    27419. .49797
                     /note="assembly_name:Contig34"
     misc_feature    49898. .79378
                     /note="assembly_name:Contig35"
     misc_feature    79479. .132595
                     /note="assembly_name:Contig36"
     misc_feature    132696. .215050
                     /note="assembly_name:Contig37"
BASE COUNT    56024 a  52513 c  52885 g  53012 t    616 others
ORIGIN
```

```
  Query Match             6.3%;  Score 92;  DB 2;  Length 215050;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-43;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205870 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 205929

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||||
Db     205930 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 205961


RESULT 11
BC036847
LOCUS       BC036847               2370 bp    mRNA    linear   PRI 10-JUN-2003
DEFINITION  Homo sapiens neurogenin 2, mRNA (cDNA clone MGC:46562
            IMAGE:5247719), complete cds.
ACCESSION   BC036847
VERSION     BC036847.1  GI:22477416
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2370)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2370)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
```

```
          Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
          Young,A., Zhang,L.-H. and Green,E.D.

          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAK Plate: 78 Row: k Column: 16.
FEATURES             Location/Qualifiers
     source          1. .2370
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:46562 IMAGE:5247719"
                     /tissue_type="Brain, fetal, whole pooled"
                     /clone_lib="NIH_MGC_121"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1. .2370
                     /gene="NEUROG2"
                     /note="synonyms: Atoh4, Math4A, ngn-2, MGC46562"
                     /db_xref="LocusID:63973"
                     /db_xref="MIM:606624"
     CDS             328. .1146
                     /codon_start=1
                     /product="neurogenin 2"
                     /protein_id="AAH36847.1"
                     /db_xref="GI:22477417"
                     /db_xref="LocusID:63973"
                     /translation="MFVKSETLELKEEEDVLVLLGSASPALAALTPLSSSADEEEEEE
                     PGASGGARRQRGAEAGQGARGGVAAGAEGCRPARLLGLVHDCKRRPSRARAVSRGAKT
                     AETVQRIKKTRRLKANNRERNRMHNLNAALDALREVLPTFPEDAKLTKIETLRFAHNY
                     IWALTETLRLADHCGGGGGGLPGALFSEAVLLSPGGASAALSSSGDSPSPASTWSCTN
                     SPAPSSSVSSNSTSPYSCTLSPASPAGSDMDYWQPPPPDKHRYAPHLPIARDCI"
BASE COUNT      580 a    654 c    616 g    520 t
ORIGIN

  Query Match             2.4%;  Score 35;  DB 9;  Length 2370;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-08;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        819 AAGATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||||||
Db        775 AAGATCGAGACCCTGCGCTTCGCCCACAACTACAT 809


RESULT 12
AF303002
LOCUS       AF303002                6123 bp    DNA     linear   PRI 13-NOV-2001
DEFINITION  Homo sapiens neurogenin 2 gene, partial cds.
ACCESSION   AF303002
VERSION     AF303002.1  GI:11875763
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6123)
  AUTHORS   Simmons,A.D., Horton,S., Abney,A.L. and Johnson,J.E.
  TITLE     Neurogenin2 expression in ventral and dorsal spinal neural tube
            progenitor cells is regulated by distinct enhancers
  JOURNAL   Dev. Biol. 229 (2), 327-339 (2001)
  MEDLINE   21077024
   PUBMED   11203697
REFERENCE   2  (bases 1 to 6123)
  AUTHORS   Simmons,A.D., Horton,S., Abney,A.L. and Johnson,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2000) Center for Basic Neuroscience - NA4.146, UT
            Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX
            75390-9111, USA
FEATURES             Location/Qualifiers
     source          1. .6123
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     mRNA            <1. .>754
                     /product="neurogenin 2"
     CDS             <1. .754
                     /note="NGN2"
                     /codon_start=2
                     /product="neurogenin 2"
                     /protein_id="AAG40770.1"
                     /db_xref="GI:11875764"
                     /translation="ASPALAALTPLSSSADEEEEEEPGASGGARRQRGAEAGQGARGG
                     VAAGAEGCRPARLLGLVHDCKRRPSRARAVSRGAKTAETVQRIKKTRRLKANNRERNR
                     MHNLNAALDALREVLPTFPEDAKLTKIETLRFAHNYIWALTETLRLADHCGGGGGGLP
                     GALFSEAVLLSPGGASAALSSSGDSPSPASTWSCTNSPAPSSSVSSNSTSPYSCTLSP
                     ASPAGSDMDYWQPPPPDKHRYAPHLPIARDCI"
BASE COUNT     1484 a   1536 c   1507 g   1596 t
ORIGIN

  Query Match             2.4%;  Score 35;  DB 9;  Length 6123;
  Best Local Similarity   100.0%;  Pred. No. 2e-08;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        819 AAGATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||||||
Db        383 AAGATCGAGACCCTGCGCTTCGCCCACAACTACAT 417


RESULT 13
AC079846_3/c
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AC079846 Accession AC079846
     Fragment Name      Begin      End
     AC079846_0             1    110000
     AC079846_1        100001    210000
     AC079846_2        200001    310000
     AC079846_3        300001    391531
Continuation (4 of 4) of AC079846 from base 300001 (AC079846 Homo sapiens chromosome 3 c

  Query Match             2.4%;  Score 35;  DB 2;  Length 91531;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-08;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        819 AAGATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||||||
Db      86995 AAGATCGAGACCCTGCGCTTCGCCCACAACTACAT 86961


RESULT 14
AC023886/c
LOCUS       AC023886              179697 bp    DNA     linear   PRI 20-MAR-2002
DEFINITION  Homo sapiens BAC clone RP11-402J6 from 4, complete sequence.
ACCESSION   AC023886
VERSION     AC023886.7  GI:19482381
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179697)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 179697)
  AUTHORS   Radionenko,M. and Abbott,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-402J6
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 179697)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 179697)
```

```
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 179697)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 15, 2002 this sequence version replaced gi:17352441.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0402J06
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is AC004049, 2000 bp overlap; the
            clone sequenced to the right is RP11-148B6.  Actual end of this
            clone is at base position 179697 of RP11-402J6.
FEATURES             Location/Qualifiers
     source          1. .179697
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-402J6"
                     /clone_lib="RPCI-11"
     repeat_region   1. .749
                     /rpt_family="L1"
     repeat_region   734. .3067
                     /rpt_family="L1"
     repeat_region   3104. .3444
                     /rpt_family="MaLR"
     repeat_region   3447. .4289
                     /rpt_family="L1"
     repeat_region   4290. .4369
                     /rpt_family="(TA)n"
     repeat_region   4370. .4797
                     /rpt_family="L1"
     repeat_region   5022. .5194
                     /rpt_family="L2"
     repeat_region   5214. .5628
                     /rpt_family="L1"
     repeat_region   5710. .5729
                     /rpt_family="(TA)n"
     repeat_region   5749. .5863
                     /rpt_family="(TA)n"
     repeat_region   5883. .6263
                     /rpt_family="ERVL"
     repeat_region   9901. .10151
                     /rpt_family="ERVL"
     repeat_region   10152. .10310
                     /rpt_family="L1"
     repeat_region   10311. .10400
                     /rpt_family="Mariner"
     repeat_region   10401. .10816
                     /rpt_family="L1"
     repeat_region   10967. .12362
                     /rpt_family="L1"
     repeat_region   12366. .12653
                     /rpt_family="Alu"
     repeat_region   12695. .13940
                     /rpt_family="L1"
     repeat_region   13965. .14272
                     /rpt_family="Alu"
     repeat_region   14274. .14795
                     /rpt_family="ERVL"
     repeat_region   14806. .15392
                     /rpt_family="L1"
     repeat_region   21504. .22112
                     /rpt_family="L1"
     repeat_region   22131. .23166
                     /rpt_family="L1"
     repeat_region   23246. .23386
                     /rpt_family="MaLR"
     repeat_region   23511. .23654
                     /rpt_family="MIR"
     repeat_region   23655. .25486
                     /rpt_family="L1"
     repeat_region   25487. .25636
                     /rpt_family="MIR"
     repeat_region   25814. .26124
                     /rpt_family="Alu"
     repeat_region   26444. .26592
                     /rpt_family="MIR"
     repeat_region   27423. .27713
                     /rpt_family="Alu"
     repeat_region   27716. .27739
                     /rpt_family="AT_rich"
     repeat_region   27773. .27795
                     /rpt_family="AT_rich"
     repeat_region   28020. .28429
                     /rpt_family="ERVL"
     repeat_region   29646. .29668
                     /rpt_family="AT_rich"
     repeat_region   29727. .29871
                     /rpt_family="L1"
     repeat_region   29875. .30114
                     /rpt_family="Alu"
     repeat_region   30134. .30267
                     /rpt_family="L1"
     repeat_region   30268. .30549
                     /rpt_family="L1"
     repeat_region   30550. .30883
                     /rpt_family="MaLR"
     repeat_region   32461. .32791
                     /rpt_family="MaLR"
     repeat_region   32792. .33487
                     /rpt_family="L1"
     repeat_region   33488. .33847
                     /rpt_family="MaLR"
```

```
     repeat_region   33848. .34354
                     /rpt_family="L1"
     repeat_region   34358. .34379
                     /rpt_family="(CATATA)n"
     repeat_region   34381. .34502
                     /rpt_family="Alu"
     repeat_region   34503. .34556
                     /rpt_family="L1"
     repeat_region   34557. .34859
                     /rpt_family="Alu"
     repeat_region   34860. .35041
                     /rpt_family="L1"
     repeat_region   35219. .35676
                     /rpt_family="L2"
     repeat_region   35974. .36042
                     /rpt_family="L2"
     repeat_region   36833. .37144
                     /rpt_family="Alu"
     misc_feature    36833. .36848
                     /note="match to EST AI537116 (NID:g4451251) to15h09.x1"
     repeat_region   37147. .37410
                     /rpt_family="Alu"
     repeat_region   38154. .38424
                     /rpt_family="Alu"
     repeat_region   39941. .40230
                     /rpt_family="Alu"
     misc_feature    39941. .39956
                     /note="match to EST AA775240 (NID:g2834574) ad18a05.s1"

  Query Match             2.4%;  Score 35;  DB 9;  Length 179697;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-08;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        819 AAGATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||||||
Db      55815 AAGATCGAGACCCTGCGCTTCGCCCACAACTACAT 55781


RESULT 15
HSA133776
LOCUS       HSA133776               1330 bp    DNA     linear   PRI 19-JUN-1999
DEFINITION  Homo sapiens gene for neurogenin 3.
ACCESSION   AJ133776
VERSION     AJ133776.1  GI:5123782
KEYWORDS    bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ravassard,P., Icard-Liepkalns,C., Wiard,L., Julien,J.P. and
            Mallet,J.
  TITLE     The human neurogenin 3 homolog maps to chromosome 10q21.3 and its
            expression pattern is identical to that of its murine counterparts
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1330)
  AUTHORS   Ravassard,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-1999) Ravassard P., Lgn, CNRS UMRC 9923, Hopital
            de la Pitie Salpetriere, Bat. CERVI, 83 Bd. de l'Hopital, 75013
            PARIS, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1330
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            1. .1330
                     /gene="ngn3"
     mRNA            join(<1. .157,321. .>1330)
                     /gene="ngn3"
     exon            1. .157
                     /gene="ngn3"
                     /number=1
     intron          158. .320
                     /gene="ngn3"
                     /number=1
     exon            321. .1330
                     /gene="ngn3"
                     /number=2
     CDS             322. .966
                     /gene="ngn3"
                     /function="early neurogenesis"
                     /codon_start=1
                     /product="neurogenin 3"
                     /protein_id="CAB45384.1"
                     /db_xref="GI:5123783"
                     /translation="MTPQPSGAPTVQVTRETERSFPRASEDEVTCPTSAPPSPTRTPG
                     NCAEAEEGGCRGAPRKLRARRGGRSRPKSELALSKQRRSRRKKANDRERNRMHDLNSA
                     LDALRGVLPTFPDDAKLTKIETLRFAHNYIWALTQTLRIADHSLYALEPPAPHCGELG
                     SPGGPPGDWGSLYSPVSQAGSLSPAASLEERPGLLGATSSACLSPGSLAFSDFL"
BASE COUNT      230 a    459 c    413 g    228 t
ORIGIN

  Query Match             2.2%;  Score 32;  DB 9;  Length 1330;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-06;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        672 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 703
              ||||||||||||||||||||||||||||||||
Db        535 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 566


Search completed: January 29, 2004, 21:29:57
Job time : 5314 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 19:49:27 ; Search time 117 Seconds
(without alignments)
5507.858 Million cell updates/sec

Title: US-09-595-947E-1
Perfect score: 1460
Sequence: 1 gcaggtagcgagaggagcag.....agagtgacctaatccagtgt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	92	6.3	804	4	US-08-932-411A-19	Sequence 19, Appl
	2	23	1.6	738	4	US-08-722-570-13	Sequence 13, Appl
	3	23	1.6	738	4	US-08-932-411A-13	Sequence 13, Appl
	4	23	1.6	1333	1	US-08-910-973-21	Sequence 21, Appl
	5	23	1.6	1333	4	US-09-499-227-21	Sequence 21, Appl
	6	23	1.6	1385	4	US-08-932-411A-17	Sequence 17, Appl
	7	21	1.4	310	1	US-08-552-142A-12	Sequence 12, Appl
	8	21	1.4	1268	1	US-08-910-973-12	Sequence 12, Appl
	9	21	1.4	1268	4	US-09-499-227-12	Sequence 12, Appl
	10	21	1.4	1352	1	US-08-552-142A-10	Sequence 10, Appl
	11	21	1.4	1535	1	US-08-910-973-10	Sequence 10, Appl
	12	21	1.4	1535	4	US-09-499-227-10	Sequence 10, Appl
	13	21	1.4	1550	3	US-09-234-332-3	Sequence 3, Appli
c	14	19	1.3	50	3	US-08-358-627F-4	Sequence 4, Appli
c	15	19	1.3	50	3	US-08-793-044-11	Sequence 11, Appl
c	16	19	1.3	50	4	US-08-465-712C-4	Sequence 4, Appli
c	17	19	1.3	50	4	US-09-552-733-4	Sequence 4, Appli
c	18	18	1.2	480	1	US-08-438-123-16	Sequence 16, Appl
	19	18	1.2	1336	4	US-09-016-434-129	Sequence 129, App
	20	18	1.2	1462	1	US-08-552-142A-16	Sequence 16, Appl
	21	18	1.2	1951	1	US-08-910-973-16	Sequence 16, Appl
	22	18	1.2	1951	4	US-09-499-227-16	Sequence 16, Appl
	23	18	1.2	3446	4	US-09-620-312D-653	Sequence 653, App
	24	18	1.2	4066	4	US-09-367-750-1	Sequence 1, Appli
	25	18	1.2	4797	4	US-09-419-568F-25	Sequence 25, Appl
	26	18	1.2	4797	4	US-09-354-243B-25	Sequence 25, Appl
c	27	18	1.2	31728	4	US-09-453-702B-64	Sequence 64, Appl
	28	17	1.2	24	3	US-08-358-627F-6	Sequence 6, Appli
	29	17	1.2	24	4	US-08-465-712C-6	Sequence 6, Appli
	30	17	1.2	24	4	US-09-552-733-6	Sequence 6, Appli
c	31	17	1.2	24	4	US-09-687-731-2	Sequence 2, Appli
c	32	17	1.2	24	4	US-09-687-731-8	Sequence 8, Appli
c	33	17	1.2	24	4	US-09-687-731-10	Sequence 10, Appl
	34	17	1.2	510	4	US-09-252-991A-9622	Sequence 9622, Ap
	35	17	1.2	685	1	US-08-751-782-5	Sequence 5, Appli
	36	17	1.2	685	2	US-08-925-171-5	Sequence 5, Appli
c	37	17	1.2	846	4	US-09-252-991A-9635	Sequence 9635, Ap
	38	17	1.2	948	4	US-09-252-991A-14810	Sequence 14810, A
	39	17	1.2	954	4	US-09-252-991A-1904	Sequence 1904, Ap
	40	17	1.2	1062	4	US-09-252-991A-9532	Sequence 9532, Ap
c	41	17	1.2	1161	1	US-08-086-439C-1	Sequence 1, Appli
c	42	17	1.2	1161	1	US-08-434-877-1	Sequence 1, Appli
c	43	17	1.2	1238	4	US-09-183-861-75	Sequence 75, Appl
c	44	17	1.2	1238	4	US-09-022-765-75	Sequence 75, Appl
c	45	17	1.2	1238	4	US-09-551-974A-75	Sequence 75, Appl

ALIGNMENTS

```
RESULT 1
US-08-932-411A-19
; Sequence 19, Application US/08932411A
; Patent No. 6566496
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr Hohbach Test Albritton & Herbert LLP
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/932,411A
;      FILING DATE:  15-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/772,009
;      FILING DATE:  19-DEC-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,570
;      FILING DATE:  19-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902-3/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  160..801
US-08-932-411A-19
```

```
  Query Match             6.3%;  Score 92;  DB 4;  Length 804;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-37;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 522

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              ||||||||||||||||||||||||||||||||
Db        523 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 554


RESULT 2
US-08-722-570-13
; Sequence 13, Application US/08722570
; Patent No. 6555337
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  5365
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-13

  Query Match             1.6%;  Score 23;  DB 4;  Length 738;
  Best Local Similarity   100.0%;  Pred. No. 0.072;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        304 GAGCGCAACCGCATGCACAACCT 326


RESULT 3
US-08-932-411A-13
; Sequence 13, Application US/08932411A
; Patent No. 6566496
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr Hohbach Test Albritton & Herbert LLP
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/932,411A
;      FILING DATE:  15-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/772,009
;      FILING DATE:  19-DEC-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,570
;      FILING DATE:  19-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902-3/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..732
US-08-932-411A-13

  Query Match             1.6%;  Score 23;  DB 4;  Length 738;
  Best Local Similarity   100.0%;  Pred. No. 0.072;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        304 GAGCGCAACCGCATGCACAACCT 326


RESULT 4
US-08-910-973-21
; Sequence 21, Application US/08910973
; Patent No. 5795723
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,973
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17532
;      FILING DATE:  30-October-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR-1-10958
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100; 206-224-0735 (direct)
;      TELEFAX:  206-225-0779
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
;    IMMEDIATE SOURCE:
;      CLONE:  neuroD3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  101..835
US-08-910-973-21

  Query Match             1.6%;  Score 23;  DB 1;  Length 1333;
  Best Local Similarity   100.0%;  Pred. No. 0.072;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        404 GAGCGCAACCGCATGCACAACCT 426


RESULT 5
US-09-499-227-21
; Sequence 21, Application US/09499227
; Patent No. 6444463
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/499,227
;      FILING DATE:  05-August-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-May-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-May-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17532
;      FILING DATE:  30-October-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/910,973
;      FILING DATE:  07-August-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR-1-12742
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100; 206-224-0735 (direct)
;      TELEFAX:  206-225-0779
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
;    IMMEDIATE SOURCE:
;      CLONE:  neuroD3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  101..835
US-09-499-227-21

  Query Match             1.6%;  Score 23;  DB 4;  Length 1333;
  Best Local Similarity   100.0%;  Pred. No. 0.072;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        404 GAGCGCAACCGCATGCACAACCT 426


RESULT 6
US-08-932-411A-17
; Sequence 17, Application US/08932411A
; Patent No. 6566496
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr Hohbach Test Albritton & Herbert LLP
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/932,411A
;      FILING DATE:  15-SEP-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/772,009
;      FILING DATE:  19-DEC-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,570
;      FILING DATE:  19-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
```

```
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902-3/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1385 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  382..1170
US-08-932-411A-17

  Query Match             1.6%;  Score 23;  DB 4;  Length 1385;
  Best Local Similarity   100.0%;  Pred. No. 0.072;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        742 GAGCGCAACCGCATGCACAACCT 764


RESULT 7
US-08-552-142A-12
; Sequence 12, Application US/08552142A
; Patent No. 5695995
;  GENERAL INFORMATION:
;    APPLICANT:  Weintraub, Harold M.
;    APPLICANT:  Lee, Jacqueline E.
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Hollenberg, Stanley M.
;    TITLE OF INVENTION:  Neurogenic Differentiation (NeuroD) Genes
;    TITLE OF INVENTION:  and Proteins
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/552,142A
;      FILING DATE:  02-NOV-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-MAY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Broderick, Thomas F.
;      REGISTRATION NUMBER:  31,332
;      REFERENCE/DOCKET NUMBER:  FHCR-1-8933
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100
;      TELEFAX:  206-225-0709
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  310 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  20A1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..310
US-08-552-142A-12

  Query Match             1.4%;  Score 21;  DB 1;  Length 310;
  Best Local Similarity   100.0%;  Pred. No. 0.72;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAAC 752
              |||||||||||||||||||||
Db        128 GAGCGCAACCGCATGCACAAC 148


RESULT 8
US-08-910-973-12
; Sequence 12, Application US/08910973
; Patent No. 5795723
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,973
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17532
;      FILING DATE:  30-October-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR-1-10958
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100; 206-224-0735 (direct)
;      TELEFAX:  206-225-0779
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1268 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  20A1 (neuroD3)
;    FEATURE:
```

```
;      NAME/KEY:  CDS
;      LOCATION:  55..768
US-08-910-973-12

  Query Match             1.4%;  Score 21;  DB 1;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 0.74;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAAC 752
              |||||||||||||||||||||
Db        355 GAGCGCAACCGCATGCACAAC 375


RESULT 9
US-09-499-227-12
; Sequence 12, Application US/09499227
; Patent No. 6444463
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/499,227
;      FILING DATE:  05-August-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-May-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-May-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17532
;      FILING DATE:  30-October-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/910,973
;      FILING DATE:  07-August-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR-1-12742
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100; 206-224-0735 (direct)
;      TELEFAX:  206-225-0779
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1268 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  20A1 (neuroD3)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  55..768
US-09-499-227-12

  Query Match             1.4%;  Score 21;  DB 4;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 0.74;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAAC 752
              |||||||||||||||||||||
Db        355 GAGCGCAACCGCATGCACAAC 375


RESULT 10
US-08-552-142A-10
; Sequence 10, Application US/08552142A
; Patent No. 5695995
;  GENERAL INFORMATION:
;    APPLICANT:  Weintraub, Harold M.
;    APPLICANT:  Lee, Jacqueline E.
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Hollenberg, Stanley M.
;    TITLE OF INVENTION:  Neurogenic Differentiation (NeuroD) Genes
;    TITLE OF INVENTION:  and Proteins
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/552,142A
;      FILING DATE:  02-NOV-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-MAY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Broderick, Thomas F.
;      REGISTRATION NUMBER:  31,332
;      REFERENCE/DOCKET NUMBER:  FHCR-1-8933
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100
;      TELEFAX:  206-225-0709
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1352 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  14B1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  55..1194
US-08-552-142A-10

  Query Match             1.4%;  Score 21;  DB 1;  Length 1352;
  Best Local Similarity   100.0%;  Pred. No. 0.74;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        729 CGGGAGCGCAACCGCATGCAC 749
              |||||||||||||||||||||
Db        439 CGGGAGCGCAACCGCATGCAC 459
```

```
RESULT 11
US-08-910-973-10
; Sequence 10, Application US/08910973
; Patent No. 5795723
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/910,973
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17532
;      FILING DATE:  30-October-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR-1-10958
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100; 206-224-0735 (direct)
;      TELEFAX:  206-225-0779
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  14B1 (neuroD2)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  55..1194
US-08-910-973-10

  Query Match             1.4%;  Score 21;  DB 1;  Length 1535;
  Best Local Similarity   100.0%;  Pred. No. 0.74;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        729 CGGGAGCGCAACCGCATGCAC 749
              |||||||||||||||||||||
Db        439 CGGGAGCGCAACCGCATGCAC 459


RESULT 12
US-09-499-227-10
; Sequence 10, Application US/09499227
; Patent No. 6444463
;  GENERAL INFORMATION:
;    APPLICANT:  Tapscott, Stephen J.
;    APPLICANT:  Olson, James M.
;    TITLE OF INVENTION:  Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson KindnessPLLC
;      STREET:  1420 Fifth Avenue, Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/499,227
;      FILING DATE:  05-August-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,238
;      FILING DATE:  06-May-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US95/05741
;      FILING DATE:  08-May-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17532
;      FILING DATE:  30-October-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/910,973
;      FILING DATE:  07-August-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR-1-12742
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100; 206-224-0735 (direct)
;      TELEFAX:  206-225-0779
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1535 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      CLONE:  14B1 (neuroD2)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  55..1194
US-09-499-227-10

  Query Match             1.4%;  Score 21;  DB 4;  Length 1535;
  Best Local Similarity   100.0%;  Pred. No. 0.74;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        729 CGGGAGCGCAACCGCATGCAC 749
              |||||||||||||||||||||
Db        439 CGGGAGCGCAACCGCATGCAC 459


RESULT 13
US-09-234-332-3
; Sequence 3, Application US/09234332A
; Patent No. 6087168
; GENERAL INFORMATION:
;  APPLICANT: Cedars-Sinai Medical Center
;  APPLICANT: Michel F. Levesque, M.D.
;  APPLICANT: Toomas Neuman, Ph.D.
;  TITLE OF INVENTION: CONVERSION OF NON-NEURONAL CELLS INTO
;  TITLE OF INVENTION: NEURONS; TRANSDIFFERENTIATION OF EPIDERMAL CELLS
;  FILE REFERENCE: P07 41494
```

```
;  CURRENT APPLICATION NUMBER: US/09/234,332A
;  CURRENT FILING DATE: 1999-01-20
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 1550
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: gene
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Neurogenic basic-helix-loop-helix protein (Neuro
;   OTHER INFORMATION: D2) gene Genbank Accession U58681
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1219)...(1226)
;   OTHER INFORMATION: n at 1219 and 1226; n = A, T, G, or C
US-09-234-332-3

  Query Match             1.4%;  Score 21;  DB 3;  Length 1550;
  Best Local Similarity   100.0%;  Pred. No. 0.74;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        729 CGGGAGCGCAACCGCATGCAC 749
              |||||||||||||||||||||
Db        442 CGGGAGCGCAACCGCATGCAC 462


RESULT 14
US-08-358-627F-4/c
; Sequence 4, Application US/08358627F
; Patent No. 6177242
;  GENERAL INFORMATION:
;    APPLICANT:  Changeux, Jean-Pierre
;    APPLICANT:  Picciotto, Marina
;    APPLICANT:  Bessis, Alain
;    TITLE OF INVENTION:  Genomic DNA Fragments Containing
;    TITLE OF INVENTION:  Regulatory and Coding Sequences For The B2-Subunit of the
;    TITLE OF INVENTION:  Neuronal Nicotinic Acetylcholine Receptor and Transgenic
;    TITLE OF INVENTION:  Animals Made Using These Fragments or Mutated Fragments
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner, L.L.P.
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/358,627F
;      FILING DATE:  14-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03495.0135-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-358-627F-4

  Query Match             1.3%;  Score 19;  DB 3;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 7.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCAGGTAGCGAGAGGAGCA 19
              |||||||||||||||||||
Db         33 GCAGGTAGCGAGAGGAGCA 15


RESULT 15
US-08-793-044-11/c
; Sequence 11, Application US/08793044
; Patent No. 6235497
;  GENERAL INFORMATION:
;    APPLICANT:  Bejanin, Stephane
;    APPLICANT:  Berrard, Sylvie
;    APPLICANT:  Cervini, Riccardo
;    APPLICANT:  Mallet, Jacques
;    TITLE OF INVENTION:  NOVEL VESICULAR ACETYLCHOLINE CARRIER
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rhone-Poulenc Rorer Inc.
;      STREET:  500 Arcola Road, Mailstop 3C43
;      CITY:  Collegeville
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19426
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/793,044
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR95/01073
;      FILING DATE:  10-AUG-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 94/10044
;      FILING DATE:  16-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Savitzky Esq., Martin F.
;      REGISTRATION NUMBER:  29,699
;      REFERENCE/DOCKET NUMBER:  ST94066-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610) 454-3816
;      TELEFAX:  (610) 454-3808
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  50 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-793-044-11

  Query Match             1.3%;  Score 19;  DB 3;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 7.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCAGGTAGCGAGAGGAGCA 19
              |||||||||||||||||||
Db         33 GCAGGTAGCGAGAGGAGCA 15


Search completed: January 29, 2004, 22:20:35
Job time : 119 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 26, 2004, 19:39:06 ; Search time 426 Seconds
(without alignments)
9251.595 Million cell updates/sec

Title: US-09-595-947E-1
Perfect score: 1460
Sequence: 1 gcaggtagcgagaggagcag.....agagtgacctaatccagtgt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1409	96.5	1491	19	AAV42512	cDNA encoding a no
2	92	6.3	804	19	AAV27050	Mouse neurogenin 3
3	92	6.3	804	21	AAZ51981	Murine neurogenin-
4	92	6.3	861	22	AAF27266	Mouse neurogenin 3
5	92	6.3	861	25	ABV75970	Mouse transcriptio
6	92	6.3	1860	24	AAD46872	Murine neurogenin
7	92	6.3	1861	21	AAC61090	Murine neurogenin
8	92	6.3	5567	22	AAF27254	Mouse atonal homol
9	36	2.5	65	24	ABN31392	Rat spliced transc
10	35	2.4	65	24	ABN57521	Mouse spliced tran
11	35	2.4	6123	24	AAD46890	Human neurogenin 2
12	32	2.2	5340	21	AAC61089	Human neurogenin 3
13	32	2.2	5340	24	AAD46871	Human neurogenin 3
14	29	2.0	428	22	AAS33797	Human cDNA encodin
c 15	26	1.8	26	24	ABT03700	Human Neurogenin-3
16	25	1.7	25	19	AAV42515	PCR primer used to
17	23	1.6	738	19	AAV27046	Mouse neurogenin 1
18	23	1.6	738	21	AAZ51977	Murine neurogenin-
19	23	1.6	790	22	AAF27264	Chicken atonal hom
20	23	1.6	1332	19	AAV42938	DNA encoding murin
21	23	1.6	1332	25	ABS56396	Mouse bHLH family
22	23	1.6	1333	18	AAT74894	Mouse neurogenic d
23	23	1.6	1385	19	AAV27049	Mouse neurogenin 2
24	23	1.6	1385	21	AAZ51980	Murine neurogenin-
25	23	1.6	1385	22	AAF27269	Mouse neurogenin 2
26	23	1.6	1412	22	AAF27255	Mouse atonal homol
27	23	1.6	1412	22	AAF27273	Mouse atonal homol
28	21	1.4	21	25	AAD47278	Human RT-PCR upstr
29	21	1.4	21	25	AAD47326	Human RT-PCR forwa
30	21	1.4	714	24	AAD46889	Human neurogenin 1
31	21	1.4	1268	18	AAT74891	Human neurogenic d
32	21	1.4	1268	19	AAV42932	DNA encoding human
33	21	1.4	1268	25	ABS56390	Human bHLH family
34	21	1.4	1535	18	AAT74890	Human neurogenic d
35	21	1.4	1535	19	AAV42931	DNA encoding human
36	21	1.4	1535	25	ABS56389	Human bHLH family
37	21	1.4	1550	21	AAA62681	Human NeuroD2 gene
38	21	1.4	1665	24	AAD46888	Human neurogenin 1
c 39	21	1.4	2776	22	AAL04043	Human reproductive
c 40	21	1.4	2776	22	AAL04045	Human reproductive
41	21	1.4	2776	22	AAK68475	Human immune/haema
42	21	1.4	2776	22	AAK68476	Human immune/haema
c 43	20	1.4	352	24	ABS69045	Novel murine polyn
c 44	20	1.4	592	24	ABQ49522	Oligonucleotide fo
45	20	1.4	592	24	ABQ49523	Oligonucleotide fo

ALIGNMENTS

RESULT 1
AAV42512
ID AAV42512 standard; cDNA; 1491 BP.
XX
AC AAV42512;
XX
DT 05-OCT-1998 (first entry)
XX
DE cDNA encoding a novel BHLH protein designated RELAX.
XX
KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;
KW control; gene expression; transcriptional activator; targeting;
KW protein expression; central nervous system; CNS; treatment;
KW nervous system disorder; ss.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT CDS 459..1103
FT /*tag= a
FT /product= RELAX
XX
PN WO9827206-A2.
XX
PD 25-JUN-1998.
XX
PF 19-DEC-1997; 97WO-FR02368.
XX
PR 19-DEC-1996; 96FR-0015651.
XX
PA (RHON) RHONE-POULENC RORER SA.

XX
PI Mallet J, Ravassard P, Icard-Liepkalns C;
XX
DR WPI; 1998-362775/31.
DR P-PSDB; AAW62991.
XX
PT Basic helix-loop-helix polypeptide and related nucleic acid - with
PT transcriptional activity, for targeting expression of genes to
PT central nervous system and treatment of nervous disease
XX
PS Claim 6; Page 20; 28pp; French.
XX
CC The present sequence encodes a basic helix-loop-helix (BHLH) type
CC protein, designated RELAX (Rat Embryonic Longitudinal Axis) protein.
CC The protein is used to control and participate in gene expression,
CC by acting as transcriptional activator, strictly dependent on the
CC presence of an intact E box (CANNTG), particularly for targeting
CC expression of proteins to the central nervous system (CNS). The
CC nucleic acid sequence can be used to treat nervous system disorders,
CC and antisense sequences can be used to control mRNA transcription.
XX
SQ Sequence 1491 BP; 307 A; 487 C; 413 G; 284 T; 0 other;

Query Match 96.5%; Score 1409; DB 19; Length 1491;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCAGGTAGCGAGAGGAGCAGTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGCA 60

Qy         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCAGCCCGGCAGGCACGCTCCTGGTCCGGGCAGAGCAGATAAAGCGTGCCAGGGGACACA 120

Qy        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGATTAGCAGCTCAGAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCCGAGGACCCCCT 180

Qy        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCGAGCTTCTTTGCTGCCTCCAGACGCAATTTACTCCAGGCGAGGGCGCCTGCAGCTCAG 240

Qy        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAAACTTCGAAGCGAGCAGAGGGGTTCAGCTATCCACCGCTGCTTGACTCTGACCACCC 300

Qy        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCAGCTCTCTGTTCTTTTGAGCCCGGAGTAACTAGGTAACATTTAGGAACCTCCAAAGGG 360

Qy        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TAGAAGAGGGGAGTGGGTGGGCGTACTCTAGTCCCGCGTGGAGTGACCTCTAAGTCAGAG 420

Qy        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTGTCACACCCCCCTTCCATTTTTTCCCAACCTCAGGATGGCGCCTCATCCCTTGGATG 480

Qy        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGCCCACCATCCAAGTGTCCCAAGAGACCCAGCAACCCTTTCCCGGAGCCTCGGACCACG 540

Qy        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCGTACCGAGGGACTGCTCCG 600

Qy        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAGCAGAAGCAGGTGACTGCCGAGGGACATCGAGGAAGCTCCGTGCGCGGCGCGGAGGGC 660

Qy        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GCAACAGGCCCAAGAGCGAGTTGGCACTGAGCAAGCAGCGACGAAGCCGGCGCAAGAAGG 720

Qy        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCAACGACCGGGAGCGCAACCGCATGCACAACCTTAACTCCGCGCTGGATGCGCTGCGCG 780

Qy        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTCG 840

Qy        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCCACAACTACATTTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTACG 900

Qy        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAGCCCGGGAGGGGGCTCCAGCGGCG 960

Qy        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACTGGGGCTCTATCTACTCCCCAGTTTCCCAAGCTGGTAGCCTGAGCCCCACAGCCTCAT 1020

Qy       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGGAGGAGTTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTGTCTGCTCCCGGGCACCC 1080

Qy       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGGTGTTCTCAGACTTCTTGTGAAGGGCCCAAACAGGCCCTGGGCGGTGGGCGCTGGCAG 1140

Qy       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAAGGGAGGGAGTCAGAGCTGTCTGAAATGGAAGGTAGTGGAGGCACTCGAGCATCTCGC 1200

Qy       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CCCTTCTGGCTTTCATTAGTCAGGTCCCTGATTTAACCAGGATTCGCACAGTTCCTTGCT 1260

Qy       1261 GCTGTGCGTGCACAAAGGATATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db       1261 GCTGTGCGTGCACAAAGGACATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGTGGCC 1320

Qy       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACCTCAAACTCCCGCTCCAAGCAGAGGAGAGCCGTAGCACTAAATAGTTGGGAGACTCCC 1380

Qy       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ATACTTCCTGGTGACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAACCACCGCTTTCTCC 1440

Qy       1441 AGAGTGACCTAATCCAGTGT 1460
              ||||||||||||||||||||
Db       1441 AGAGTGACCTAATCCAGTGT 1460
```

RESULT 2
AAV27050
ID AAV27050 standard; cDNA; 804 BP.
XX
AC AAV27050;
XX
DT 17-SEP-1998 (first entry)
XX
DE Mouse neurogenin 3 gene.
XX
KW ds; Mouse; neurogenin; expression vector; recombinant protein;
KW antibody; neurogenesis.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 160..804
FT /*tag= a

```
FT                   /product= "Mouse neurogenin 3"
XX
PN   WO9813491-A2.
XX
PD   02-APR-1998.
XX
PF   24-SEP-1997;   97WO-US17048.
XX
PR   17-SEP-1997;   97US-0932411.
PR   27-SEP-1996;   96US-0722570.
PR   12-NOV-1996;   96US-0030864.
PR   19-DEC-1996;   96US-0772009.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Anderson DJ,  Ma Q,  Sommer L;
XX
DR   WPI; 1998-230702/20.
DR   P-PSDB; AAW54947.
XX
PT   Mouse neurogenins, useful in neurogenesis - and recombinant nucleic
PT   acids and proteins derived from rat and Xenopus
XX
PS   Disclosure; Fig 9; 106pp; English.
XX
CC   The Mouse neurogenin 3 is one of several neurogenin proteins discussed
CC   in the present invention.  The neurogenin nucleic acids can be expressed
CC   in a host cell, transformed using an expression vector, to produce
CC   recombinant proteins.  The proteins and the antibodies raised against
CC   the proteins are useful in the study of neurogenesis.
XX
SQ   Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

  Query Match             6.3%;  Score 92;  DB 19;  Length 804;
  Best Local Similarity   100.0%;  Pred. No. 9e-35;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 522

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||
Db        523 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 554


RESULT 3
AAZ51981
ID   AAZ51981 standard; DNA; 804 BP.
XX
AC   AAZ51981;
XX
DT   04-JUL-2000  (first entry)
XX
DE   Murine neurogenin-3 (NGN3) nucleic acid sequence.
XX
KW   Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;
KW   Phox2a protein; neuronal subtype-specific marker; growth factor;
KW   neural differentiation; transplantation; neuronal dysfunction;
KW   optical nerve damage; auditory nerve damage, neurodegenerative disorder;
KW   neuroprotective; nootropic; anticonvulsant; antiParkinsonian; vulnerary;
KW   cerebroprotective; immunesuppressant; antiinfectious; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             160..804
FT                   /*tag=  a
FT                   /product= "Murine neurogenin-3 protein"
XX
PN   WO200009676-A2.
XX
PD   24-FEB-2000.
XX
PF   13-AUG-1999;   99WO-US18525.
XX
PR   14-AUG-1998;   98US-0096630.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Anderson DJ,  Lo L;
XX
DR   WPI; 2000-256250/22.
DR   P-PSDB; AAY70570.
XX
PT   Inducing non-neuronal cells to differentiate into neurons and for
PT   non-neuronal cells to express a neuronal subtype-specific marker,
PT   comprising contacting the non-neuronal cells with a vector containing
PT   neurogenin nucleic acid -
XX
PS   Claim 1; Fig 1J; 76pp; English.
XX
CC   The patent discloses a method for inducing non-neuronal cells (NNC) to
CC   differentiate into neurons and for NNCs to express a neuronal subtype
CC   -specific marker. Transformed host cells are used as sources of neuronal
CC   and other growth factors; in culture for screening compounds that
CC   modulate neural differentiation or as sources of recombinantly produced
CC   neurogenins and Phox2a proteins for use in transplantation. The cells
CC   also have a variety of in vivo uses, e.g. for transplantation at sites of
CC   neuronal dysfunction e.g. patients with hearing or vision loss due to
CC   optical or auditory nerve damage, brain or spinal cord injuries, and
CC   neurodegenerative disorders e.g. Alzheimer's disease. The present
CC   sequence encodes murine neurogenin-3 (NGN-3), a transcription factor.
CC   NNCs differentiate into neurons through the recombinant expression of a
CC   transcription factor that induces a core program of neurogenesis. Forced
CC   expression of murine NGN3 can elicit expression of at least some neuronal
CC   phenotypic markers even in NNCs.
XX
SQ   Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

  Query Match             6.3%;  Score 92;  DB 21;  Length 804;
  Best Local Similarity   100.0%;  Pred. No. 9e-35;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 522

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||
Db        523 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 554


RESULT 4
AAF27266
ID   AAF27266 standard; cDNA; 861 BP.
XX
AC   AAF27266;
XX
DT   24-APR-2001  (first entry)
XX
DE   Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.
XX
KW   Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW   hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW   cellular proliferation; cerebellar granule neuron; gene therapy;
KW   mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW   transgenic animal; ss.
XX
OS   Mus musculus.
XX
PN   WO200073764-A2.
XX
PD   07-DEC-2000.
XX
PF   01-JUN-2000; 2000WO-US15410.
```

XX
PR 01-JUN-1999; 99US-0137060.
PR 19-JAN-2000; 2000US-0176993.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
XX
DR WPI; 2001-032190/04.
DR P-PSDB; AAB60359.
XX
PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT any of its homologs or orthologs, for the treatment of e.g. deafness,
PT osteoarthritis and abnormal cell proliferation -
XX
PS Disclosure; Page -; 142pp; English.
XX
CC The invention relates to the use of atonal-associated nucleic acid or
CC amino acid sequence, or any of its homologues or orthologues as
CC therapeutic agents for the treatment of deafness, partial hearing loss,
CC vestibular effects due to damage or loss of inner hair cells,
CC osteoarthritis and abnormal cell proliferation. The invention also
CC encompasses methods of screening for compounds which affect the
CC expression of an atonal-associated nucleic acid sequence in an animal,
CC and a transgenic animal in which an allele of a native atonal-associated
CC gene is replaced by a heterologous nucleic acid sequence, thus
CC inactivating the atonal-associated allele. The nucleic acids or proteins
CC may be used in a method of treating an animal for hearing impairment,
CC joint disease, balance disorders, abnormal cell proliferation, or other
CC disease related to loss of a functional atonal-associated nucleic acid or
CC protein. They may particularly be used to treat an animal with a
CC deficiency in cerebellar granule neurons or their precursors, and may
CC also be used in promoting mechanoreceptive cell growth and generating
CC hair cells. The present sequence represents an atonal-associated nucleic
CC acid sequence referred to in the invention.
CC Note: The present sequence is not shown in the specification, but
CC was obtained from GenBank.
XX
SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

Query Match 6.3%; Score 92; DB 22; Length 861;
Best Local Similarity 100.0%; Pred. No. 9e-35;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 522

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||
Db        523 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 554
```

RESULT 5
ABV75970
ID ABV75970 standard; cDNA; 861 BP.
XX
AC ABV75970;
XX
DT 11-FEB-2003 (first entry)
XX
DE Mouse transcription factor neurogenin 3 cDNA.
XX
KW Mouse; transcription factor; neurogenin 3; ngn3; stem cell;
KW differentiation; beta-cell; insulin; diabetes; hyperglycaemia;
KW glucose intolerance; antidiabetic; hypoglycaemic; gene therapy;
KW gene; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 160..804
FT /*tag= a
FT /product= "Mouse ngn3"
XX
PN WO200286107-A2.
XX
PD 31-OCT-2002.
XX
PF 19-APR-2002; 2002WO-EP04362.
XX
PR 19-APR-2001; 2001US-284531P.
XX
PA (DEVE-) DEVELOGEN ENTWICKLUNGSBIOLOGISCHE FORSCH.
PA (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.
XX
PI Wobus AM, St-Onge L, Blyszczuk P, Hoffmann U;
XX
DR WPI; 2003-075629/07.
XX
PT Differentiating stem cells into insulin-producing cells useful for
PT treating pancreatic diseases, by culturing stem cells in suitable
PT medium and activating gene involved in beta-cell differentiation -
XX
PS Disclosure; Page 58-59; 62pp; English.
XX
CC The present sequence is that of cDNA encoding the murine
CC basic helix-loop-helix transcription factor neurogenin 3 (ngn3),
CC a gene which is required for the specification of the early
CC endocrine precursor in the pancreatic epithelium and which is
CC down-regulated once endocrine differentiation begins. The invention
CC provides a claimed method for differentiating stem cells (especially
CC embryonic, adult or somatic stem cells and primordial germ cells)
CC into insulin-producing cells. This involves culturing stem cells in
CC a suitable medium and activating at least one gene involved in
CC beta-cell differentiation. Preferred genes including Pdx1, Pax4,
CC Pax6 and ngn3 (see ABV75967-70). Gene activation comprises the
CC delivery of the gene into stem cells using a viral delivery
CC system, or the delivery of a protein product of the gene into stem
CC cells. The insulin-producing cells can be transplanted into
CC animals or human for treatment of pancreatic diseases, metabolic
CC syndrome and metabolic disorders with impaired glucose levels such
CC as diabetes, hyperglycaemia and impaired glucose tolerance
CC (claimed). The cells can also be used to identify compounds which
CC stimulate beta-cell differentiation, insulin secretion or glucose
CC responsiveness. Differentiated beta-cells can be used to study the
CC toxic and other effects of exogenous compounds on beta-cell
CC function. In an example from the invention, Pax6 cDNA was inserted
CC into expression vector pACCMV.pLpA under the control of the
CC cytomegalovirus promoter.
XX
SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

Query Match 6.3%; Score 92; DB 25; Length 861;
Best Local Similarity 100.0%; Pred. No. 9e-35;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 522

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||
Db        523 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 554
```

RESULT 6
AAD46872
ID AAD46872 standard; DNA; 1860 BP.
XX
AC AAD46872;
XX
DT 27-JAN-2003 (first entry)
XX
DE Murine neurogenin 3 (Ngn3) gene.
XX

```
KW   Transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW   type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW   islet cell; cell therapy; neurogenin 3; Ngn3; murine; gene; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1093..1737
FT                   /*tag=  a
FT                   /product= "Murine Ngn3 protein"
XX
PN   WO200274045-A2.
XX
PD   26-SEP-2002.
XX
PF   20-MAR-2002; 2002WO-US11166.
XX
PR   20-MAR-2001; 2001US-0817360.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   German MS,  Lin J;
XX
DR   WPI; 2002-759853/82.
DR   P-PSDB; AAE29278.
XX
PT   Producing a mammalian islet cell for treating diabetes mellitus
PT   comprises introducing into a mammalian cell a nucleic acid molecule
PT   encoding neuroendocrine basic helix-loop-helix transcription factor  -
XX
PS   Example 3; Page 89-90; 108pp; English.
XX
CC   The invention relates to a method for producing a mammalian islet cell.
CC   The method comprising introducing into a mammalian cell a nucleic acid
CC   molecule encoding an islet transcription factor for expression of the
CC   islet transcription factor in the cell and for production of islet cell
CC   phenotype in the cell. The islet transcription factor is a neuroendocrine
CC   basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC   for treating type 2 diabetes mellitus and for replacing beta cells lost
CC   to autoimmune destruction in individuals with type 1 diabetes. The method
CC   is useful in cell therapy. The present sequence is murine neurogenin 3
CC   (Ngn3) gene.
XX
SQ   Sequence 1860 BP; 397 A; 559 C; 537 G; 367 T; 0 other;

  Query Match             6.3%;  Score 92;  DB 24;  Length 1860;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-35;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1396 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 1455

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||
Db       1456 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 1487


RESULT 7
AAC61090
ID   AAC61090 standard; DNA; 1861 BP.
XX
AC   AAC61090;
XX
DT   05-FEB-2001  (first entry)
XX
DE   Murine neurogenin 3 (Ngn3) genomic DNA sequence.
XX
KW   Neurogenin 3; Ngn3; cellular differentiation; diabetes mellitus;
KW   islet cell precursor identification; mouse; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1093..1737
FT                   /*tag=  a
FT                   /product= "Ngn3"
FT                   /note= "Neurogenin 3"
XX
PN   WO200059936-A1.
XX
PD   12-OCT-2000.
XX
PF   28-MAR-2000; 2000WO-US08436.
XX
PR   06-APR-1999;   99US-0128180.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   German MS,  Lin J;
XX
DR   WPI; 2000-664989/64.
DR   P-PSDB; AAY85618.
XX
PT   Novel human neurogenin 3 polypeptides and polynucleotides encoding
PT   them, useful for diagnosis, prevention and treatment of diabetes
PT   mellitus and to identify individuals at risk of diabetes -
XX
PS   Claim 18; Page 49-50; 54pp; English.
XX
CC   The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
CC   protein AAY85617. The Ngn3 gene is located at chromosome position
CC   10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
CC   protein sequences, and includes an antibody recognising the Ngn3 protein.
CC   Also included in the invention is a method for identifying an islet cell
CC   precursor, the method involves analysing a cell for the expression of the
CC   Ngn3 gene product, where detection of the product is indicative of an
CC   islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
CC   reagent for detecting (in a subject) a predisposition to a defect in
CC   pancreatic islet cell function or formation associated with a defect in
CC   Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
CC   precursor cells expressing Ngn3, and to alter cellular differentiation in
CC   culture in vivo to produce new beta-cells to treat patients with diabetes
CC   mellitus. The present sequence represents the murine Ngn3 genomic DNA
CC   sequence.
XX
SQ   Sequence 1861 BP; 397 A; 560 C; 537 G; 367 T; 0 other;

  Query Match             6.3%;  Score 92;  DB 21;  Length 1861;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-35;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1396 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 1455

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||
Db       1456 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 1487


RESULT 8
AAF27254
ID   AAF27254 standard; cDNA; 5567 BP.
XX
AC   AAF27254;
XX
DT   24-APR-2001  (first entry)
XX
DE   Mouse atonal homologue 5 (ATOH5, Math4B) cDNA, SEQ ID NO:4.
XX
KW   Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW   hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW   cellular proliferation; cerebellar granule neuron; gene therapy;
KW   mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW   transgenic animal; ss.
```

XX
OS Mus musculus.
XX
PN WO200073764-A2.
XX
PD 07-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15410.
XX
PR 01-JUN-1999; 99US-0137060.
PR 19-JAN-2000; 2000US-0176993.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
XX
DR WPI; 2001-032190/04.
DR P-PSDB; AAB60350.
XX
PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT any of its homologs or orthologs, for the treatment of e.g. deafness,
PT osteoarthritis and abnormal cell proliferation -
XX
PS Disclosure; Page -; 142pp; English.
XX
CC The invention relates to the use of atonal-associated nucleic acid or
CC amino acid sequence, or any of its homologues or orthologues as
CC therapeutic agents for the treatment of deafness, partial hearing loss,
CC vestibular effects due to damage or loss of inner hair cells,
CC osteoarthritis and abnormal cell proliferation. The invention also
CC encompasses methods of screening for compounds which affect the
CC expression of an atonal-associated nucleic acid sequence in an animal,
CC and a transgenic animal in which an allele of a native atonal-associated
CC gene is replaced by a heterologous nucleic acid sequence, thus
CC inactivating the atonal-associated allele. The nucleic acids or proteins
CC may be used in a method of treating an animal for hearing impairment,
CC joint disease, balance disorders, abnormal cell proliferation, or other
CC disease related to loss of a functional atonal-associated nucleic acid or
CC protein. They may particularly be used to treat an animal with a
CC deficiency in cerebellar granule neurons or their precursors, and may
CC also be used in promoting mechanoreceptive cell growth and generating
CC hair cells. The present sequence represents an atonal-associated nucleic
CC acid sequence referred to in the invention.
CC Note: The present sequence is not shown in the specification, but
CC was obtained from GenBank.
XX
SQ Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;

Query Match 6.3%; Score 92; DB 22; Length 5567;
Best Local Similarity 100.0%; Pred. No. 8.3e-35;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 5226 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 5285

Qy 822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
|||||||||||||||||||||||||||||||
Db 5286 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 5317

RESULT 9
ABN31392
ID ABN31392 standard; DNA; 65 BP.
XX
AC ABN31392;
XX
DT 15-JUL-2002 (first entry)
XX
DE Rat spliced transcript detection oligonucleotide SEQ ID NO:4140.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Rattus norvegicus.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-IB01903.
XX
PR 28-JUL-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes -
XX
PS Example 1; SEQ ID 4140; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridising selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 65 BP; 11 A; 20 C; 14 G; 20 T; 0 other;

Query Match 2.5%; Score 36; DB 24; Length 65;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1281 ATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGT 1316
||||||||||||||||||||||||||||||||||||
Db 30 ATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGT 65

RESULT 10
ABN57521
ID ABN57521 standard; DNA; 65 BP.
XX
AC ABN57521;
XX
DT 15-JUL-2002 (first entry)
XX
DE Mouse spliced transcript detection oligonucleotide SEQ ID NO:30269.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.

XX
OS Mus musculus.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-IB01903.
XX
PR 28-JUL-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes -
XX
PS Example 1; SEQ ID 30269; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridising selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 65 BP; 16 A; 25 C; 11 G; 13 T; 0 other;

Query Match 2.4%; Score 35; DB 24; Length 65;
Best Local Similarity 100.0%; Pred. No. 9.8e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        547 TCAGTTCCAATTCCACCCCACCTAGCCCCACTCTC 581
              |||||||||||||||||||||||||||||||||||
Db          1 TCAGTTCCAATTCCACCCCACCTAGCCCCACTCTC 35
```

RESULT 11
AAD46890
ID AAD46890 standard; DNA; 6123 BP.
XX
AC AAD46890;
XX
DT 27-JAN-2003 (first entry)
XX
DE Human neurogenin 2 (Ngn2) gene #1.
XX
KW Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW islet cell; cell therapy; neurogenin 2; Ngn2; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 2..754
FT /*tag= a
FT /product= "Human Ngn2 protein"
FT /note= "No start codon"
FT /partial
XX
PN WO200274045-A2.
XX
PD 26-SEP-2002.
XX
PF 20-MAR-2002; 2002WO-US11166.
XX
PR 20-MAR-2001; 2001US-0817360.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI German MS, Lin J;
XX
DR WPI; 2002-759853/82.
DR P-PSDB; AAE29281.
XX
PT Producing a mammalian islet cell for treating diabetes mellitus
PT comprises introducing into a mammalian cell a nucleic acid molecule
PT encoding neuroendocrine basic helix-loop-helix transcription factor -
XX
PS Disclosure; Page 95-97; 108pp; English.
XX
CC The invention relates to a method for producing a mammalian islet cell.
CC The method comprising introducing into a mammalian cell a nucleic acid
CC molecule encoding an islet transcription factor for expression of the
CC islet transcription factor in the cell and for production of islet cell
CC phenotype in the cell. The islet transcription factor is a neuroendocrine
CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC for treating type 2 diabetes mellitus and for replacing beta cells lost
CC to autoimmune destruction in individuals with type 1 diabetes. The method
CC is useful in cell therapy. The present sequence is human neurogenin 2
CC (Ngn2) gene.
XX
SQ Sequence 6123 BP; 1484 A; 1536 C; 1507 G; 1596 T; 0 other;

Query Match 2.4%; Score 35; DB 24; Length 6123;
Best Local Similarity 100.0%; Pred. No. 8.1e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        819 AAGATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||||||
Db        383 AAGATCGAGACCCTGCGCTTCGCCCACAACTACAT 417
```

RESULT 12
AAC61089
ID AAC61089 standard; DNA; 5340 BP.
XX
AC AAC61089;
XX
DT 05-FEB-2001 (first entry)
XX
DE Human neurogenin 3 (Ngn3) genomic DNA sequence.
XX
KW Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;
KW islet cell precursor identification; diabetes mellitus; human; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 3022..3666
FT /*tag= a
FT /product= "Ngn3"

```
FT                   /note= "Neurogenin 3"
XX
PN   WO200059936-A1.
XX
PD   12-OCT-2000.
XX
PF   28-MAR-2000; 2000WO-US08436.
XX
PR   06-APR-1999;   99US-0128180.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   German MS,  Lin J;
XX
DR   WPI; 2000-664989/64.
DR   P-PSDB; AAY85617.
XX
PT   Novel human neurogenin 3 polypeptides and polynucleotides encoding
PT   them, useful for diagnosis, prevention and treatment of diabetes
PT   mellitus and to identify individuals at risk of diabetes -
XX
PS   Claim 6; Page 46-48; 54pp; English.
XX
CC   The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
CC   protein AAY85617. The Ngn3 gene is located at chromosome position
CC   10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
CC   protein sequences, and includes an antibody recognising the Ngn3 protein.
CC   Also included in the invention is a method for identifying an islet cell
CC   precursor, the method involves analysing a cell for the expression of the
CC   Ngn3 gene product, where detection of the product is indicative of an
CC   islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
CC   reagent for detecting (in a subject) a predisposition to a defect in
CC   pancreatic islet cell function or formation associated with a defect in
CC   Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
CC   precursor cells expressing Ngn3, and to alter cellular differentiation in
CC   culture in vivo to produce new beta-cells to treat patients with diabetes
CC   mellitus.
XX
SQ   Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

  Query Match             2.2%;  Score 32;  DB 21;  Length 5340;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-05;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        672 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 703
              ||||||||||||||||||||||||||||||||
Db       3235 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 3266


RESULT 13
AAD46871
ID   AAD46871 standard; DNA; 5340 BP.
XX
AC   AAD46871;
XX
DT   27-JAN-2003  (first entry)
XX
DE   Human neurogenin 3 (Ngn3) gene.
XX
KW   Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW   type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW   islet cell; cell therapy; neurogenin 3; Ngn3; chromosome 10q22.1-22.2;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             3022..3666
FT                   /*tag=  a
FT                   /product= "Human Ngn3 protein"
XX
PN   WO200274045-A2.
XX
PD   26-SEP-2002.
XX
PF   20-MAR-2002; 2002WO-US11166.
XX
PR   20-MAR-2001; 2001US-0817360.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   German MS,  Lin J;
XX
DR   WPI; 2002-759853/82.
DR   P-PSDB; AAE29277.
XX
PT   Producing a mammalian islet cell for treating diabetes mellitus
PT   comprises introducing into a mammalian cell a nucleic acid molecule
PT   encoding neuroendocrine basic helix-loop-helix transcription factor  -
XX
PS   Example 2; Page 87-88; 108pp; English.
XX
CC   The invention relates to a method for producing a mammalian islet cell.
CC   The method comprising introducing into a mammalian cell a nucleic acid
CC   molecule encoding an islet transcription factor for expression of the
CC   islet transcription factor in the cell and for production of islet cell
CC   phenotype in the cell. The islet transcription factor is a neuroendocrine
CC   basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC   for treating type 2 diabetes mellitus and for replacing beta cells lost
CC   to autoimmune destruction in individuals with type 1 diabetes. The method
CC   is useful in cell therapy. The present sequence is human neurogenin 3
CC   (Ngn3) gene. Ngn3 gene is located on chromosome 10q22.1-22.2.
XX
SQ   Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

  Query Match             2.2%;  Score 32;  DB 24;  Length 5340;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-05;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        672 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 703
              ||||||||||||||||||||||||||||||||
Db       3235 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 3266


RESULT 14
AAS33797
ID   AAS33797 standard; cDNA; 428 BP.
XX
AC   AAS33797;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human cDNA encoding a novel foetal antigen, SEQ ID No 321.
XX
KW   Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;
KW   immunomodulator; cardiovascular; cytostatic; nephrothropic;
KW   cardiovascular; autoimmune disease; rheumatoid arthritis;
KW   hyperproliferative disorder; breast neoplasm; cancer;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200155312-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01321.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
```

PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488782/53.
DR P-PSDB; AAU20977.
XX
PT New polynucleotides and polypeptides for diagnosing, treating,
PT preventing or prognosing e.g. diseases or disorders of the nervous,
PT musculoskeletal, excretory, gastrointestinal, reproductive, and
PT respiratory systems -
XX
PS Claim 1; SEQ ID No 321; 642pp; English.
XX

CC The invention relates to novel nucleic acids encoding novel human foetal
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. The antibodies to the antigens can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. Numerous
CC examples of diseases and disorders treated by the nucleic acids and
CC proteins are given in the specification. The present sequence

Query Match 2.0%; Score 29; DB 22; Length 428;
Best Local Similarity 100.0%; Pred. No. 0.0008;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 825 GAGACCCTGCGCTTCGCCCACAACTACAT 853
||||||||||||||||||||||||||||||
Db 1 GAGACCCTGCGCTTCGCCCACAACTACAT 29

RESULT 15
ABT03700/c
ID ABT03700 standard; DNA; 26 BP.
XX
AC ABT03700;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human Neurogenin-3 gene PCR primer SEQ ID NO: 221.
XX
KW Human; cancer; neoplastic disease; tumour specific marker; cytostatic;
KW transcription factor; PCR; primer; ss.
XX
OS Homo sapiens.
XX
PN WO200240716-A2.
XX
PD 23-MAY-2002.
XX
PF 13-NOV-2001; 2001WO-US43461.
XX
PR 16-NOV-2000; 2000US-249508P.
XX
PA (CEMI-) CEMINES LLC.
XX
PI Palm K;
XX
DR WPI; 2002-537346/57.
XX
PT Determining the presence of neoplastic molecular markers, by
PT identifying the presence of markers in host test sample using array of
PT neoplastic molecular marker specific reagents and analyzing the array
PT of the reagents -
XX
PS Example 1; Page 17; 41pp; English.
XX
CC The present invention relates to a method for determining the presence of
CC neoplastic molecular markers in a host, involving the use of neoplastic
CC molecular marker specific reagents to detect such markers and analysing
CC the array of reagents, allowing the identification of the neoplastic
CC disease present. This can be used to determine the best treatment for
CC cancers, in particular neural cell, lung and prostate tumours. The
CC present sequence is a PCR primer useful for detecting the coding
CC sequences of markers of the invention.
XX
SQ Sequence 26 BP; 3 A; 10 C; 5 G; 8 T; 0 other;

Query Match 1.8%; Score 26; DB 24; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 673 AGAGCGAGTTGGCACTGAGCAAGCAG 698
||||||||||||||||||||||||||
Db 26 AGAGCGAGTTGGCACTGAGCAAGCAG 1

RESULT 16
AAV42515
ID AAV42515 standard; DNA; 25 BP.
XX
AC AAV42515;
XX
DT 05-OCT-1998 (first entry)
XX
DE PCR primer used to isolate part of the RELAX protein coding region.
XX
KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;
KW control; gene expression; transcriptional activator; targeting;
KW protein expression; central nervous system; CNS; treatment;
KW nervous system disorder; CIG235; PCR primer; ss.
XX
OS Synthetic.
OS Rattus sp.
XX
PN WO9827206-A2.
XX
PD 25-JUN-1998.
XX
PF 19-DEC-1997; 97WO-FR02368.
XX
PR 19-DEC-1996; 96FR-0015651.
XX
PA (RHON) RHONE-POULENC RORER SA.
XX
PI Mallet J, Ravassard P, Icard-Liepkalns C;
XX
DR WPI; 1998-362775/31.
XX
PT Basic helix-loop-helix polypeptide and related nucleic acid - with
PT transcriptional activity, for targeting expression of genes to
PT central nervous system and treatment of nervous disease
XX
PS Example 2; Page 12; 28pp; French.
XX
CC PCR primers AAV42515-16 are used to isolate part of the DNA encoding
CC a basic helix-loop-helix (BHLH) type protein, designated RELAX (Rat
CC Embryonic Longitudinal Axis) protein. The PCR product is termed CIG235.
CC The protein is used to control and participate in gene expression,
CC by acting as transcriptional activator, strictly dependent on the
CC presence of an intact E box (CANNTG), particularly for targeting
CC expression of proteins to the central nervous system (CNS). The
CC nucleic acid sequence can be used to treat nervous system disorders,
CC and antisense sequences can be used to control mRNA transcription.
XX
SQ Sequence 25 BP; 5 A; 9 C; 6 G; 5 T; 0 other;

Query Match 1.7%; Score 25; DB 19; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.083;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 750 AACCTTAACTCCGCGCTGGATGCGC 774
|||||||||||||||||||||||||

```
Db          1 AACCTTAACTCCGCGCTGGATGCGC 25


RESULT 17
AAV27046
ID   AAV27046 standard; cDNA; 738 BP.
XX
AC   AAV27046;
XX
DT   17-SEP-1998  (first entry)
XX
DE   Mouse neurogenin 1 gene.
XX
KW   ss; Mouse; neurogenin; expression vector; recombinant protein;
KW   antibody; neurogenesis.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..735
FT                   /*tag=  a
FT                   /product= "Mouse neurogenin 1"
XX
PN   WO9813491-A2.
XX
PD   02-APR-1998.
XX
PF   24-SEP-1997;   97WO-US17048.
XX
PR   17-SEP-1997;   97US-0932411.
PR   27-SEP-1996;   96US-0722570.
PR   12-NOV-1996;   96US-0030864.
PR   19-DEC-1996;   96US-0772009.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Anderson DJ,  Ma Q,  Sommer L;
XX
DR   WPI; 1998-230702/20.
DR   P-PSDB; AAW54944.
XX
PT   Mouse neurogenins, useful in neurogenesis - and recombinant nucleic
PT   acids and proteins derived from rat and Xenopus
XX
PS   Claim 5; Fig 4; 106pp; English.
XX
CC   The mouse neurogenin 1 is one of several neurogenin proteins discussed in
CC   the present invention.  The neurogenin nucleic acids can be expressed in
CC   a host cell, transformed using an expression vector, to produce
CC   recombinant proteins.  The proteins and the antibodies raised against
CC   the proteins are useful in the study of neurogenesis.
XX
SQ   Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;

  Query Match             1.6%;  Score 23;  DB 19;  Length 738;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        304 GAGCGCAACCGCATGCACAACCT 326


RESULT 18
AAZ51977
ID   AAZ51977 standard; DNA; 738 BP.
XX
AC   AAZ51977;
XX
DT   04-JUL-2000  (first entry)
XX
DE   Murine neurogenin-1 (NGN1) nucleic acid sequence.
XX
KW   Neurogenin-1; NGN-1; non-neuronal cell; NNC; neurogenesis;
KW   Phox2a protein; neuronal subtype-specific marker; growth factor;
KW   neural differentiation; transplantation; neuronal dysfunction;
KW   optical nerve damage; auditory nerve damage, neurodegenerative disorder;
KW   neuroprotective; nootropic; anticonvulsant; antiParkinsonian; vulnerary;
KW   cerebroprotective; immunesuppressant; antiinfectious; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..735
FT                   /*tag=  a
FT                   /product= "Murine neurogenin-1 protein"
XX
PN   WO200009676-A2.
XX
PD   24-FEB-2000.
XX
PF   13-AUG-1999;   99WO-US18525.
XX
PR   14-AUG-1998;   98US-0096630.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Anderson DJ,  Lo L;
XX
DR   WPI; 2000-256250/22.
DR   P-PSDB; AAY70566.
XX
PT   Inducing non-neuronal cells to differentiate into neurons and for
PT   non-neuronal cells to express a neuronal subtype-specific marker,
PT   comprising contacting the non-neuronal cells with a vector containing
PT   neurogenin nucleic acid -
XX
PS   Claim 1; Fig 1C; 76pp; English.
XX
CC   The patent discloses a method for inducing non-neuronal cells (NNC) to
CC   differentiate into neurons and for NNCs to express a neuronal subtype
CC   -specific marker. Transformed host cells are used as sources of neuronal
CC   and other growth factors; in culture for screening compounds that
CC   modulate neural differentiation or as sources of recombinantly produced
CC   neurogenins and Phox2a proteins for use in transplantation. The cells
CC   also have a variety of in vivo uses, e.g. for transplantation at sites of
CC   neuronal dysfunction e.g. patients with hearing or vision loss due to
CC   optical or auditory nerve damage, brain or spinal cord injuries, and
CC   neurodegenerative disorders e.g. Alzheimer's disease. The present
CC   sequence encodes murine neurogenin-1 (NGN-1), a transcription factor.
CC   NNCs differentiate into neurons through the recombinant expression of a
CC   transcription factor that induces a core program of neurogenesis. Forced
CC   expression of murine NGN1 can elicit expression of at least some neuronal
CC   phenotypic markers even in NNCs. This can be used in autografting.
XX
SQ   Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;

  Query Match             1.6%;  Score 23;  DB 21;  Length 738;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        304 GAGCGCAACCGCATGCACAACCT 326


RESULT 19
AAF27264
ID   AAF27264 standard; cDNA; 790 BP.
XX
AC   AAF27264;
XX
DT   24-APR-2001  (first entry)
XX
DE   Chicken atonal homologue ngn2/ath4a cDNA, SEQ ID NO:20.
XX
```

KW Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW cellular proliferation; cerebellar granule neuron; gene therapy;
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW transgenic animal; ss.
XX
OS Gallus gallus.
XX
PN WO200073764-A2.
XX
PD 07-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15410.
XX
PR 01-JUN-1999; 99US-0137060.
PR 19-JAN-2000; 2000US-0176993.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
XX
DR WPI; 2001-032190/04.
DR P-PSDB; AAB60357.
XX
PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT any of its homologs or orthologs, for the treatment of e.g. deafness,
PT osteoarthritis and abnormal cell proliferation -
XX
PS Disclosure; Page -; 142pp; English.
XX
CC The invention relates to the use of atonal-associated nucleic acid or
CC amino acid sequence, or any of its homologues or orthologues as
CC therapeutic agents for the treatment of deafness, partial hearing loss,
CC vestibular effects due to damage or loss of inner hair cells,
CC osteoarthritis and abnormal cell proliferation. The invention also
CC encompasses methods of screening for compounds which affect the
CC expression of an atonal-associated nucleic acid sequence in an animal,
CC and a transgenic animal in which an allele of a native atonal-associated
CC gene is replaced by a heterologous nucleic acid sequence, thus
CC inactivating the atonal-associated allele. The nucleic acids or proteins
CC may be used in a method of treating an animal for hearing impairment,
CC joint disease, balance disorders, abnormal cell proliferation, or other
CC disease related to loss of a functional atonal-associated nucleic acid or
CC protein. They may particularly be used to treat an animal with a
CC deficiency in cerebellar granule neurons or their precursors, and may
CC also be used in promoting mechanoreceptive cell growth and generating
CC hair cells. The present sequence represents an atonal-associated nucleic
CC acid sequence referred to in the invention.
CC Note: The present sequence is not shown in the specification, but
CC was obtained from GenBank.
XX
SQ Sequence 790 BP; 91 A; 351 C; 283 G; 65 T; 0 other;

Query Match 1.6%; Score 23; DB 22; Length 790;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACCT 754
||||||||||||||||||||||
Db 374 GAGCGCAACCGCATGCACAACCT 396

RESULT 20
AAV42938
ID AAV42938 standard; DNA; 1332 BP.
XX
AC AAV42938;
XX
DT 25-MAR-2003 (updated)
DT 21-OCT-1998 (first entry)
XX
DE DNA encoding murine neuroD3 protein, which is a bHLH protein.
XX
KW Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;
KW classification; medulloblastoma; mouse; ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 101..835
FT /*tag= a
FT /product= neuroD3
XX
PN US5795723-A.
XX
PD 18-AUG-1998.
XX
PF 07-AUG-1997; 97US-0910973.
XX
PR 06-MAY-1994; 94US-0239238.
PR 02-NOV-1995; 95US-0552142.
PR 30-OCT-1996; 96WO-US17532.
XX
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI Olson JM, Tapscott SJ;
XX
DR WPI; 1998-466661/40.
DR P-PSDB; AAW71019.
XX
PT Classifying neuroectodermal tumours from expression pattern of
PT basic-helix-loop-helix genes - especially for identifying
PT medulla:blastoma and assessing its aggressiveness, specifically
PT associated with expression of BHLH genes neuroD 1-3
XX
PS Example 11; Columns 75-78; 43pp; English.
XX
CC The present sequence encodes a protein which is a member of the basic
CC helix-loop-helix (bHLH) protein family, and is designated neuroD3. The
CC bHLH genes are a family of genes associated with vertebrate neuronal,
CC endocrinal and gastrointestinal development. The observed pattern of
CC neuroD expression distinguishes subclasses of neuroectodermal tumours.
CC The specification describes a method for the classification of human
CC neuroectodermal tumours. The method comprises measuring, in a tumour
CC sample, expression of at least one basic bHLH gene and identifying the
CC tumour subclass by matching expression to predetermined expression
CC profiles for known subclasses. For classifying the tumour as a
CC medulloblastoma, the bHLH gene detected is neuroD1 and neuroD3.
CC The method is used to classify neuroectodermal tumours, and to identify
CC medullablastoma and for prognosis of this as aggressive.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 1332 BP; 268 A; 452 C; 352 G; 260 T; 0 other;

Query Match 1.6%; Score 23; DB 19; Length 1332;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACCT 754
||||||||||||||||||||||
Db 404 GAGCGCAACCGCATGCACAACCT 426

RESULT 21
ABS56396
ID ABS56396 standard; DNA; 1332 BP.
XX
AC ABS56396;
XX
DT 23-JAN-2003 (first entry)
XX
DE Mouse bHLH family neuroD3 genomic DNA.
XX
KW Mouse; gene; ds; neuroD3; neuroD; basic-helix-loop-helix; bHLH;
KW differentiation; neurone; endocrine; gastrointestine; development;
KW transgenic; embryo; birth defect; spontaneous abortion; stem cell;

```
KW   cancer; neural growth factor; tumour; diagnostic; motor; sensory;
KW   traumatic neural injury; hearing; vision; brain; spinal cord;
KW   malabsorption syndrome; gastrointestinal dysmotility syndrome;
KW   Hirsh Prung's disease; therapeutic.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             101..835
FT                   /*tag=  a
FT                   /product= "NeuroD3"
FT   misc_feature    425..544
FT                   /*tag=  b
FT                   /note= "HLH coding domain"
XX
PN   US6444463-B1.
XX
PD   03-SEP-2002.
XX
PF   07-FEB-2000; 2000US-0499227.
XX
PR   05-AUG-1998;   98WO-US16417.
PR   06-MAY-1994;   94US-0239238.
PR   08-MAY-1995;   95WO-US05741.
PR   02-NOV-1995;   95US-0552142.
PR   30-OCT-1996;   96WO-US17532.
PR   07-AUG-1997;   97US-0910973.
XX
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI   Tapscott SJ;
XX
DR   WPI; 2003-056678/05.
DR   P-PSDB; ABG72005.
XX
PT   New neurogenic differentiation gene, useful in gene therapy to correct
PT   traumatic neural injury that has resulted in loss of motor or sensory
PT   neural function and for constructing recombinant cell lines  -
XX
PS   Example 11; Column 75-78; 43pp; English.
XX
CC   The invention discloses an isolated nucleic acid molecule which encodes a
CC   functionally active human neuroD3 polypeptide. NeuroD proteins represent
CC   a new family within the basic-helix-loop-helix (bHLH) family which are
CC   implicated in the regulation of differentiation. NeuroD proteins are
CC   particularly involved in neuronal, endocrine and gastrointestinal
CC   development. The nucleic acid is useful for constructing recombinant cell
CC   lines, transgenic embryos and animals and for quantifying the level of
CC   expression of neuroD in a cell. Birth defects and spontaneous abortions
CC   may result from expression of an abnormal neuroD protein. The
CC   polynucleotide sequences permit the establishment of primary cultures of
CC   proliferating embryonic neuronal stem cells under conditions mimicking
CC   those that are active in development and cancer. The resultant cell lines
CC   find use as sources of novel neural growth factors, in assays for
CC   identifying novel neuronal growth factors which can be used for screening
CC   anti-cancer drugs capable of driving terminal differentiation in neural
CC   tumours, for producing antibodies useful in diagnostic assays and for
CC   screening for compounds capable of modulating the activity of neuroD.
CC   Transformed host cells, nucleic acids and polypeptides are also useful
CC   for treating sites of traumatic neural injury where motor or sensory
CC   neural activity has been lost, e.g. hearing or vision loss and brain or
CC   spinal cord damage. The host cells find use in the treatment of
CC   malabsorption syndromes or gastrointestinal dysmotility syndromes (Hirsh
CC   Prung's Disease). The cell lines also find use in screening for candidate
CC   therapeutic agents capable of either substituting for neuroD or
CC   correcting the cellular defect caused by a defective neuroD. The sequence
CC   presented is the mouse neuroD3 genomic DNA.
XX
SQ   Sequence 1332 BP; 268 A; 452 C; 352 G; 260 T; 0 other;

  Query Match             1.6%;  Score 23;  DB 25;  Length 1332;
  Best Local Similarity   100.0%;  Pred. No. 0.67;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        404 GAGCGCAACCGCATGCACAACCT 426


RESULT 22
AAT74894
ID   AAT74894 standard; cDNA; 1333 BP.
XX
AC   AAT74894;
XX
DT   02-OCT-1997  (first entry)
XX
DE   Mouse neurogenic differentiation protein (NeuroD3) DNA.
XX
KW   Neurogenic differentiation protein; NeuroD; neuroD3 gene;
KW   transcriptional activator; neuron; pancreas; gastrointestine;
KW   knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;
KW   ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             101..835
FT                   /*tag=  a
XX
PN   WO9716548-A1.
XX
PD   09-MAY-1997.
XX
PF   30-OCT-1996;   96WO-US17532.
XX
PR   02-NOV-1995;   95US-0552142.
XX
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA   (WEIN/) WEINTRAUB N.
XX
PI   Hollenberg SM,  Lee JE,  Tapscott SJ,  Weintraub HM;
XX
DR   WPI; 1997-272117/24.
DR   P-PSDB; AAW22443.
XX
PT   Nucleic acid encoding neurogenic differentiation polypeptide -
PT   useful e.g. in regulating neuronal, endocrine and gastrointestinal
PT   development
XX
PS   Claim 1; Page 72-74; 81pp; English.
XX
CC   Neurogenic differentiation (NeuroD) genes (AAT74887-94) and proteins
CC   (AAW22436-43) from human, mouse and frog have been identified,
CC   isolated and sequence.  NeuroD polypeptides are tissue-specific
CC   basic-helix-loop-helix (bHLH) transcriptional activators involved
CC   in neuronal, endocrine and gastrointestinal development.  They were
CC   discovered by expression cloning and screening assays designed to
CC   identify possible bHLH proteins capable of interacting with the
CC   protein product of the Drosophila daugherless gene.  Novel neuroD2
CC   and neuroD3 genes, related to neuroD1, have been identified.
CC   NeuroD nucleic acids can be used to produce NeuroD polypeptides,
CC   construction of test cell lines, as probes, in gene therapy, and to
CC   produce transgenic animals as models of disease.
XX
SQ   Sequence 1333 BP; 268 A; 452 C; 353 G; 260 T; 0 other;

  Query Match             1.6%;  Score 23;  DB 18;  Length 1333;
  Best Local Similarity   100.0%;  Pred. No. 0.67;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        404 GAGCGCAACCGCATGCACAACCT 426
```

```
RESULT 23
AAV27049
ID   AAV27049 standard; cDNA; 1385 BP.
XX
AC   AAV27049;
XX
DT   17-SEP-1998  (first entry)
XX
DE   Mouse neurogenin 2 gene.
XX
KW   ds; Mouse; neurogenin; expression vector; recombinant protein;
KW   antibody; neurogenesis.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             382..1173
FT                   /*tag=  a
FT                   /product= "Mouse neurogenin 2"
XX
PN   WO9813491-A2.
XX
PD   02-APR-1998.
XX
PF   24-SEP-1997;   97WO-US17048.
XX
PR   17-SEP-1997;   97US-0932411.
PR   27-SEP-1996;   96US-0722570.
PR   12-NOV-1996;   96US-0030864.
PR   19-DEC-1996;   96US-0772009.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Anderson DJ,  Ma Q,  Sommer L;
XX
DR   WPI; 1998-230702/20.
DR   P-PSDB; AAW54946.
XX
PT   Mouse neurogenins, useful in neurogenesis - and recombinant nucleic
PT   acids and proteins derived from rat and Xenopus
XX
PS   Claim 5; Fig 7; 106pp; English.
XX
CC   The Mouse neurogenin 2 is one of several neurogenin proteins discussed
CC   in the present invention.  The neurogenin nucleic acids can be expressed
CC   in a host cell, transformed using an expression vector, to produce
CC   recombinant proteins.  The proteins and the antibodies raised against
CC   the proteins are useful in the study of neurogenesis.
XX
SQ   Sequence 1385 BP; 242 A; 467 C; 432 G; 244 T; 0 other;

  Query Match             1.6%;  Score 23;  DB 19;  Length 1385;
  Best Local Similarity   100.0%;  Pred. No. 0.67;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        742 GAGCGCAACCGCATGCACAACCT 764


RESULT 24
AAZ51980
ID   AAZ51980 standard; DNA; 1385 BP.
XX
AC   AAZ51980;
XX
DT   04-JUL-2000  (first entry)
XX
DE   Murine neurogenin-2 (NGN-2) nucleic acid sequence.
XX
KW   Neurogenin-2; NGN-2; non-neuronal cell; NNC; neurogenesis;
KW   Phox2a protein; neuronal subtype-specific marker; growth factor;
KW   neural differentiation; transplantation; neuronal dysfunction;
KW   optical nerve damage; auditory nerve damage, neurodegenerative disorder;
KW   neuroprotective; nootropic; anticonvulsant; antiParkinsonian; vulnerary;
KW   cerebroprotective; immunesuppressant; antiinfectious; ds.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             382..1173
FT                   /*tag=  a
FT                   /product= "Murine neurogenin-2 protein"
XX
PN   WO200009676-A2.
XX
PD   24-FEB-2000.
XX
PF   13-AUG-1999;   99WO-US18525.
XX
PR   14-AUG-1998;   98US-0096630.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Anderson DJ,  Lo L;
XX
DR   WPI; 2000-256250/22.
DR   P-PSDB; AAY70569.
XX
PT   Inducing non-neuronal cells to differentiate into neurons and for
PT   non-neuronal cells to express a neuronal subtype-specific marker,
PT   comprising contacting the non-neuronal cells with a vector containing
PT   neurogenin nucleic acid -
XX
PS   Claim 1; Fig 1I; 76pp; English.
XX
CC   The patent discloses a method for inducing non-neuronal cells (NNC) to
CC   differentiate into neurons and for NNCs to express a neuronal subtype
CC   -specific marker. Transformed host cells are used as sources of neuronal
CC   and other growth factors; in culture for screening compounds that
CC   modulate neural differentiation or as sources of recombinantly produced
CC   neurogenins and Phox2a proteins for use in transplantation. The cells
CC   also have a variety of in vivo uses, e.g. for transplantation at sites of
CC   neuronal dysfunction e.g. patients with hearing or vision loss due to
CC   optical or auditory nerve damage, brain or spinal cord injuries, and
CC   neurodegenerative disorders e.g. Alzheimer's disease. The present
CC   sequence encodes murine neurogenin-2 (NGN-2), a transcription factor.
CC   NNCs differentiate into neurons through the recombinant expression of a
CC   transcription factor that induces a core program of neurogenesis. Forced
CC   expression of murine NGN2 can elicit expression of at least some neuronal
CC   phenotypic markers even in NNCs.
XX
SQ   Sequence 1385 BP; 242 A; 467 C; 432 G; 244 T; 0 other;

  Query Match             1.6%;  Score 23;  DB 21;  Length 1385;
  Best Local Similarity   100.0%;  Pred. No. 0.67;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        742 GAGCGCAACCGCATGCACAACCT 764


RESULT 25
AAF27269
ID   AAF27269 standard; cDNA; 1385 BP.
XX
AC   AAF27269;
XX
DT   24-APR-2001  (first entry)
XX
DE   Mouse neurogenin 2 (ngn2) cDNA, SEQ ID NO:30.
XX
KW   Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW   hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW   cellular proliferation; cerebellar granule neuron; gene therapy;
```

KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW transgenic animal; ss.
XX
OS Mus musculus.
XX
PN WO200073764-A2.
XX
PD 07-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15410.
XX
PR 01-JUN-1999; 99US-0137060.
PR 19-JAN-2000; 2000US-0176993.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
XX
DR WPI; 2001-032190/04.
DR P-PSDB; AAB60362.
XX
PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT any of its homologs or orthologs, for the treatment of e.g. deafness,
PT osteoarthritis and abnormal cell proliferation -
XX
PS Disclosure; Page -; 142pp; English.
XX
CC The invention relates to the use of atonal-associated nucleic acid or
CC amino acid sequence, or any of its homologues or orthologues as
CC therapeutic agents for the treatment of deafness, partial hearing loss,
CC vestibular effects due to damage or loss of inner hair cells,
CC osteoarthritis and abnormal cell proliferation. The invention also
CC encompasses methods of screening for compounds which affect the
CC expression of an atonal-associated nucleic acid sequence in an animal,
CC and a transgenic animal in which an allele of a native atonal-associated
CC gene is replaced by a heterologous nucleic acid sequence, thus
CC inactivating the atonal-associated allele. The nucleic acids or proteins
CC may be used in a method of treating an animal for hearing impairment,
CC joint disease, balance disorders, abnormal cell proliferation, or other
CC disease related to loss of a functional atonal-associated nucleic acid or
CC protein. They may particularly be used to treat an animal with a
CC deficiency in cerebellar granule neurons or their precursors, and may
CC also be used in promoting mechanoreceptive cell growth and generating
CC hair cells. The present sequence represents an atonal-associated nucleic
CC acid sequence referred to in the invention.
CC Note: The present sequence is not shown in the specification, but
CC was obtained from GenBank.
XX
SQ Sequence 1385 BP; 242 A; 467 C; 432 G; 244 T; 0 other;

Query Match 1.6%; Score 23; DB 22; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACCT 754
||||||||||||||||||||||
Db 742 GAGCGCAACCGCATGCACAACCT 764

RESULT 26
AAF27255
ID AAF27255 standard; cDNA; 1412 BP.
XX
AC AAF27255;
XX
DT 24-APR-2001 (first entry)
XX
DE Mouse atonal homologue 4 (ATOH4, Math4A) cDNA, SEQ ID NO:6.
XX
KW Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW cellular proliferation; cerebellar granule neuron; gene therapy;
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW transgenic animal; ss.
XX
OS Mus musculus.
XX
PN WO200073764-A2.
XX
PD 07-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15410.
XX
PR 01-JUN-1999; 99US-0137060.
PR 19-JAN-2000; 2000US-0176993.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
XX
DR WPI; 2001-032190/04.
DR P-PSDB; AAB60351.
XX
PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT any of its homologs or orthologs, for the treatment of e.g. deafness,
PT osteoarthritis and abnormal cell proliferation -
XX
PS Disclosure; Page -; 142pp; English.
XX
CC The invention relates to the use of atonal-associated nucleic acid or
CC amino acid sequence, or any of its homologues or orthologues as
CC therapeutic agents for the treatment of deafness, partial hearing loss,
CC vestibular effects due to damage or loss of inner hair cells,
CC osteoarthritis and abnormal cell proliferation. The invention also
CC encompasses methods of screening for compounds which affect the
CC expression of an atonal-associated nucleic acid sequence in an animal,
CC and a transgenic animal in which an allele of a native atonal-associated
CC gene is replaced by a heterologous nucleic acid sequence, thus
CC inactivating the atonal-associated allele. The nucleic acids or proteins
CC may be used in a method of treating an animal for hearing impairment,
CC joint disease, balance disorders, abnormal cell proliferation, or other
CC disease related to loss of a functional atonal-associated nucleic acid or
CC protein. They may particularly be used to treat an animal with a
CC deficiency in cerebellar granule neurons or their precursors, and may
CC also be used in promoting mechanoreceptive cell growth and generating
CC hair cells. The present sequence represents an atonal-associated nucleic
CC acid sequence referred to in the invention.
CC Note: The present sequence is not shown in the specification, but
CC was obtained from GenBank.
XX
SQ Sequence 1412 BP; 265 A; 447 C; 435 G; 265 T; 0 other;

Query Match 1.6%; Score 23; DB 22; Length 1412;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACCT 754
||||||||||||||||||||||
Db 445 GAGCGCAACCGCATGCACAACCT 467

RESULT 27
AAF27273
ID AAF27273 standard; cDNA; 1412 BP.
XX
AC AAF27273;
XX
DT 24-APR-2001 (first entry)
XX
DE Mouse atonal homologue 4A (Math4A) cDNA, SEQ ID NO:37.
XX
KW Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW cellular proliferation; cerebellar granule neuron; gene therapy;
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW transgenic animal; ss.

XX
OS Mus musculus.
XX
PN WO200073764-A2.
XX
PD 07-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US15410.
XX
PR 01-JUN-1999; 99US-0137060.
PR 19-JAN-2000; 2000US-0176993.
XX
PA (BAYU) BAYLOR COLLEGE MEDICINE.
XX
PI Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
XX
DR WPI; 2001-032190/04.
DR P-PSDB; AAB60365.
XX
PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT any of its homologs or orthologs, for the treatment of e.g. deafness,
PT osteoarthritis and abnormal cell proliferation -
XX
PS Disclosure; Page -; 142pp; English.
XX
CC The invention relates to the use of atonal-associated nucleic acid or
CC amino acid sequence, or any of its homologues or orthologues as
CC therapeutic agents for the treatment of deafness, partial hearing loss,
CC vestibular effects due to damage or loss of inner hair cells,
CC osteoarthritis and abnormal cell proliferation. The invention also
CC encompasses methods of screening for compounds which affect the
CC expression of an atonal-associated nucleic acid sequence in an animal,
CC and a transgenic animal in which an allele of a native atonal-associated
CC gene is replaced by a heterologous nucleic acid sequence, thus
CC inactivating the atonal-associated allele. The nucleic acids or proteins
CC may be used in a method of treating an animal for hearing impairment,
CC joint disease, balance disorders, abnormal cell proliferation, or other
CC disease related to loss of a functional atonal-associated nucleic acid or
CC protein. They may particularly be used to treat an animal with a
CC deficiency in cerebellar granule neurons or their precursors, and may
CC also be used in promoting mechanoreceptive cell growth and generating
CC hair cells. The present sequence represents an atonal-associated nucleic
CC acid sequence referred to in the invention.
CC Note: The present sequence is not shown in the specification, but
CC was obtained from GenBank.
XX
SQ Sequence 1412 BP; 265 A; 447 C; 435 G; 265 T; 0 other;

Query Match 1.6%; Score 23; DB 22; Length 1412;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACCT 754
||||||||||||||||||||||
Db 445 GAGCGCAACCGCATGCACAACCT 467

RESULT 28
AAD47278
ID AAD47278 standard; DNA; 21 BP.
XX
AC AAD47278;
XX
DT 24-FEB-2003 (first entry)
XX
DE Human RT-PCR upstream primer for neurogenin-3 DNA isolation.
XX
KW Human; insulin-secreting cell; neurogenin 3; ngn3; precursor stem cell;
KW pancreatic exocrine cell; transplantation; RT-PCR; primer; ss.
XX
OS Homo sapiens.
XX
PN WO200274946-A2.
XX
PD 26-SEP-2002.
XX
PF 26-FEB-2002; 2002WO-DK00130.
XX
PR 26-FEB-2001; 2001US-271474P.
XX
PA (NOVO) NOVO NORDISK AS.
XX
PI Serup P, Heimberg H, Gradwohl G;
XX
DR WPI; 2003-018804/01.
XX
PT Generating insulin-secreting cells from precursor stem cells or adult
PT pancreatic exocrine cells, for generating glucose sensitive insulin
PT secreting beta cells for transplantation, comprises using neurogenin3
PT or NeuroD/beta2 -
XX
PS Example 4; Page 28; 66pp; English.
XX
CC The invention relates to a method for generating insulin-secreting cells
CC from precursor stem cells or adult pancreatic exocrine cells. The method
CC comprises exposing the precursor cells or exocrine cells to: a nucleic
CC acid molecule encoding neurogenin 3 (ngn3) or NeuroD/beta2; or an
CC activator of ngn3 or NeuroD/beta2 gene expression, under conditions
CC effective to generate the insulin-generating cells from the precursor or
CC exocrine cells. The invention is useful in generating insulin-secreting
CC cells from precursor stem cells or adult pancreatic exocrine cells is
CC useful for generating glucose sensitive insulin secreting beta cells
CC suitable for transplantation, and for in situ development of insulin-
CC secreting cells in a patient. The method is also useful for preventing
CC premature differentiation of precursor stem cells into insulin-secreting
CC beta cells and for identifying compounds that prevent or activate beta
CC cell differentiation. The present sequence is human RT-PCR primer for
CC isolation of neurogenin-3 DNA.
XX
SQ Sequence 21 BP; 2 A; 7 C; 6 G; 6 T; 0 other;

Query Match 1.4%; Score 21; DB 25; Length 21;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 460 TGGCGCCTCATCCCTTGGATG 480
|||||||||||||||||||||
Db 1 TGGCGCCTCATCCCTTGGATG 21

RESULT 29
AAD47326
ID AAD47326 standard; DNA; 21 BP.
XX
AC AAD47326;
XX
DT 24-FEB-2003 (first entry)
XX
DE Human RT-PCR forward primer for mouse ngn3 DNA isolation.
XX
KW Human; insulin-secreting cell; neurogenin 3; ngn3; precursor stem cell;
KW pancreatic exocrine cell; transplantation; RT-PCR; primer; ss.
XX
OS Homo sapiens.
XX
PN WO200274946-A2.
XX
PD 26-SEP-2002.
XX
PF 26-FEB-2002; 2002WO-DK00130.
XX
PR 26-FEB-2001; 2001US-271474P.
XX
PA (NOVO) NOVO NORDISK AS.
XX
PI Serup P, Heimberg H, Gradwohl G;

```
XX
DR   WPI; 2003-018804/01.
XX
PT   Generating insulin-secreting cells from precursor stem cells or adult
PT   pancreatic exocrine cells, for generating glucose sensitive insulin
PT   secreting beta cells for transplantation, comprises using neurogenin3
PT   or NeuroD/beta2  -
XX
PS   Example 5B; Page 37; 66pp; English.
XX
CC   The invention relates to a method for generating insulin-secreting cells
CC   from precursor stem cells or adult pancreatic exocrine cells. The method
CC   comprises exposing the precursor cells or exocrine cells to: a nucleic
CC   acid molecule encoding neurogenin 3 (ngn3) or NeuroD/beta2; or an
CC   activator of ngn3 or NeuroD/beta2 gene expression, under conditions
CC   effective to generate the insulin-generating cells from the precursor or
CC   exocrine cells. The invention is useful in generating insulin-secreting
CC   cells from precursor stem cells or adult pancreatic exocrine cells is
CC   useful for generating glucose sensitive insulin secreting beta cells
CC   suitable for transplantation, and for in situ development of insulin-
CC   secreting cells in a patient. The method is also useful for preventing
CC   premature differentiation of precursor stem cells into insulin-secreting
CC   beta cells and for identifying compounds that prevent or activate beta
CC   cell differentiation. The present sequence is human RT-PCR primer for
CC   isolation of mouse ngn3 DNA.
XX
SQ   Sequence 21 BP; 7 A; 7 C; 4 G; 3 T; 0 other;

  Query Match             1.4%;  Score 21;  DB 25;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 7.7;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        798 CCGGATGACGCCAAACTTACA 818
              |||||||||||||||||||||
Db          1 CCGGATGACGCCAAACTTACA 21


RESULT 30
AAD46889
ID   AAD46889 standard; DNA; 714 BP.
XX
AC   AAD46889;
XX
DT   27-JAN-2003  (first entry)
XX
DE   Human neurogenin 1 (Ngn1) gene #2.
XX
KW   Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW   type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW   islet cell; cell therapy; neurogenin 1; Ngn1; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..714
FT                   /*tag=  a
FT                   /product= "Human Ngn1 protein"
XX
PN   WO200274045-A2.
XX
PD   26-SEP-2002.
XX
PF   20-MAR-2002; 2002WO-US11166.
XX
PR   20-MAR-2001; 2001US-0817360.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   German MS,  Lin J;
XX
DR   WPI; 2002-759853/82.
DR   P-PSDB; AAE29280.
XX
PT   Producing a mammalian islet cell for treating diabetes mellitus
PT   comprises introducing into a mammalian cell a nucleic acid molecule
PT   encoding neuroendocrine basic helix-loop-helix transcription factor  -
XX
PS   Disclosure; Page 94; 108pp; English.
XX
CC   The invention relates to a method for producing a mammalian islet cell.
CC   The method comprising introducing into a mammalian cell a nucleic acid
CC   molecule encoding an islet transcription factor for expression of the
CC   islet transcription factor in the cell and for production of islet cell
CC   phenotype in the cell. The islet transcription factor is a neuroendocrine
CC   basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC   for treating type 2 diabetes mellitus and for replacing beta cells lost
CC   to autoimmune destruction in individuals with type 1 diabetes. The method
CC   is useful in cell therapy. The present sequence is human neurogenin 1
CC   (Ngn1) gene.
XX
SQ   Sequence 714 BP; 118 A; 287 C; 207 G; 102 T; 0 other;

  Query Match             1.4%;  Score 21;  DB 24;  Length 714;
  Best Local Similarity   100.0%;  Pred. No. 6.6;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAAC 752
              |||||||||||||||||||||
Db        301 GAGCGCAACCGCATGCACAAC 321


RESULT 31
AAT74891
ID   AAT74891 standard; DNA; 1268 BP.
XX
AC   AAT74891;
XX
DT   02-OCT-1997  (first entry)
XX
DE   Human neurogenic differentiation protein (NeuroD3) DNA clone 20A1.
XX
KW   Neurogenic differentiation protein; NeuroD; neuroD3 gene;
KW   transcriptional activator; neuron; pancreas; gastrointestine;
KW   knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             55..768
FT                   /*tag=  a
XX
PN   WO9716548-A1.
XX
PD   09-MAY-1997.
XX
PF   30-OCT-1996;   96WO-US17532.
XX
PR   02-NOV-1995;   95US-0552142.
XX
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA   (WEIN/) WEINTRAUB N.
XX
PI   Hollenberg SM,  Lee JE,  Tapscott SJ,  Weintraub HM;
XX
DR   WPI; 1997-272117/24.
DR   P-PSDB; AAW22440.
XX
PT   Nucleic acid encoding neurogenic differentiation polypeptide -
PT   useful e.g. in regulating neuronal, endocrine and gastrointestinal
PT   development
XX
PS   Claim 1; Page 64-65; 81pp; English.
XX
CC   Neurogenic differentiation (NeuroD) genes (AAT74887-94) and proteins
CC   (AAW22436-43) from human, mouse and frog have been identified,
```

```
CC   isolated and sequence.  NeuroD polypeptides are tissue-specific
CC   basic-helix-loop-helix (bHLH) transcriptional activators involved
CC   in neuronal, endocrine and gastrointestinal development.  They were
CC   discovered by expression cloning and screening assays designed to
CC   identify possible bHLH proteins capable of interacting with the
CC   protein product of the Drosophila daugherless gene.  Novel neuroD2
CC   and neuroD3 genes, related to neuroD1, have been identified.
CC   NeuroD nucleic acids can be used to produce NeuroD polypeptides,
CC   construction of test cell lines, as probes, in gene therapy, and to
CC   produce transgenic animals as models of disease.
XX
SQ   Sequence 1268 BP; 245 A; 455 C; 344 G; 224 T; 0 other;

  Query Match             1.4%;  Score 21;  DB 18;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 6.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAAC 752
              |||||||||||||||||||||
Db        355 GAGCGCAACCGCATGCACAAC 375


RESULT 32
AAV42932
ID   AAV42932 standard; DNA; 1268 BP.
XX
AC   AAV42932;
XX
DT   25-MAR-2003  (updated)
DT   21-OCT-1998  (first entry)
XX
DE   DNA encoding human neuroD3 protein, which is a bHLH protein.
XX
KW   Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;
KW   classification; medulloblastoma; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             55..768
FT                   /*tag=  a
FT                   /product= neuroD3
XX
PN   US5795723-A.
XX
PD   18-AUG-1998.
XX
PF   07-AUG-1997;   97US-0910973.
XX
PR   06-MAY-1994;   94US-0239238.
PR   02-NOV-1995;   95US-0552142.
PR   30-OCT-1996;   96WO-US17532.
XX
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI   Olson JM,  Tapscott SJ;
XX
DR   WPI; 1998-466661/40.
DR   P-PSDB; AAW71016.
XX
PT   Classifying neuroectodermal tumours from expression pattern of
PT   basic-helix-loop-helix genes - especially for identifying
PT   medulla:blastoma and assessing its aggressiveness, specifically
PT   associated with expression of BHLH genes neuroD 1-3
XX
PS   Example 11; Columns 61-64; 43pp; English.
XX
CC   The present sequence encodes a protein which is a member of the basic
CC   helix-loop-helix (bHLH) protein family, and is designated neuroD3. The
CC   bHLH genes are a family of genes associated with vertebrate neuronal,
CC   endocrinal and gastrointestinal development. The observed pattern of
CC   neuroD expression distinguishes subclasses of neuroectodermal tumours.
CC   The specification describes a method for the classification of human
CC   neuroectodermal tumours. The method comprises measuring, in a tumour
CC   sample, expression of at least one basic bHLH gene and identifying the
CC   tumour subclass by matching expression to predetermined expression
CC   profiles for known subclasses. For classifying the tumour as a
CC   medulloblastoma, the bHLH gene detected is neuroD1 and neuroD3.
CC   The method is used to classify neuroectodermal tumours, and to identify
CC   medullablastoma and for prognosis of this as aggressive.
CC   (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ   Sequence 1268 BP; 245 A; 455 C; 344 G; 224 T; 0 other;

  Query Match             1.4%;  Score 21;  DB 19;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 6.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAAC 752
              |||||||||||||||||||||
Db        355 GAGCGCAACCGCATGCACAAC 375


RESULT 33
ABS56390
ID   ABS56390 standard; DNA; 1268 BP.
XX
AC   ABS56390;
XX
DT   23-JAN-2003  (first entry)
XX
DE   Human bHLH family neuroD3 genomic DNA, clone 20A1.
XX
KW   Human; gene; ds; neuroD3; neuroD; basic-helix-loop-helix; bHLH;
KW   differentiation; neurone; endocrine; gastrointestine; development;
KW   transgenic; embryo; birth defect; spontaneous abortion; stem cell;
KW   cancer; neural growth factor; tumour; diagnostic; motor; sensory;
KW   traumatic neural injury; hearing; vision; brain; spinal cord;
KW   malabsorption syndrome; gastrointestinal dysmotility syndrome;
KW   Hirsh Prung's disease; therapeutic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             55..768
FT                   /*tag=  a
FT                   /product= "NeuroD3"
FT   misc_feature    376..495
FT                   /*tag=  b
FT                   /note= "HLH coding domain"
XX
PN   US6444463-B1.
XX
PD   03-SEP-2002.
XX
PF   07-FEB-2000; 2000US-0499227.
XX
PR   05-AUG-1998;   98WO-US16417.
PR   06-MAY-1994;   94US-0239238.
PR   08-MAY-1995;   95WO-US05741.
PR   02-NOV-1995;   95US-0552142.
PR   30-OCT-1996;   96WO-US17532.
PR   07-AUG-1997;   97US-0910973.
XX
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI   Tapscott SJ;
XX
DR   WPI; 2003-056678/05.
DR   P-PSDB; ABG72002.
XX
PT   New neurogenic differentiation gene, useful in gene therapy to correct
PT   traumatic neural injury that has resulted in loss of motor or sensory
PT   neural function and for constructing recombinant cell lines  -
XX
PS   Claim 1; Column 61-64; 43pp; English.
```

```
XX
CC   The invention discloses an isolated nucleic acid molecule which encodes a
CC   functionally active human neuroD3 polypeptide. NeuroD proteins represent
CC   a new family within the basic-helix-loop-helix (bHLH) family which are
CC   implicated in the regulation of differentiation. NeuroD proteins are
CC   particularly involved in neuronal, endocrine and gastrointestinal
CC   development. The nucleic acid is useful for constructing recombinant cell
CC   lines, transgenic embryos and animals and for quantifying the level of
CC   expression of neuroD in a cell. Birth defects and spontaneous abortions
CC   may result from expression of an abnormal neuroD protein. The
CC   polynucleotide sequences permit the establishment of primary cultures of
CC   proliferating embryonic neuronal stem cells under conditions mimicking
CC   those that are active in development and cancer. The resultant cell lines
CC   find use as sources of novel neural growth factors, in assays for
CC   identifying novel neuronal growth factors which can be used for screening
CC   anti-cancer drugs capable of driving terminal differentiation in neural
CC   tumours, for producing antibodies useful in diagnostic assays and for
CC   screening for compounds capable of modulating the activity of neuroD.
CC   Transformed host cells, nucleic acids and polypeptides are also useful
CC   for treating sites of traumatic neural injury where motor or sensory
CC   neural activity has been lost, e.g. hearing or vision loss and brain or
CC   spinal cord damage. The host cells find use in the treatment of
CC   malabsorption syndromes or gastrointestinal dysmotility syndromes (Hirsh
CC   Prung's Disease). The cell lines also find use in screening for candidate
CC   therapeutic agents capable of either substituting for neuroD or
CC   correcting the cellular defect caused by a defective neuroD. The sequence
CC   presented is the human neuroD3 genomic DNA, clone 20A1.
XX
SQ   Sequence 1268 BP; 245 A; 455 C; 344 G; 224 T; 0 other;

  Query Match             1.4%;  Score 21;  DB 25;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 6.5;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAAC 752
              |||||||||||||||||||||
Db        355 GAGCGCAACCGCATGCACAAC 375


RESULT 34
AAT74890
ID   AAT74890 standard; DNA; 1535 BP.
XX
AC   AAT74890;
XX
DT   02-OCT-1997  (first entry)
XX
DE   Human neurogenic differentiation protein (NeuroD2) DNA clone 14B1.
XX
KW   Neurogenic differentiation protein; NeuroD; neuroD2 gene;
KW   transcriptional activator; neuron; pancreas; gastrointestine;
KW   knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             55..1200
FT                   /*tag=  a
XX
PN   WO9716548-A1.
XX
PD   09-MAY-1997.
XX
PF   30-OCT-1996;   96WO-US17532.
XX
PR   02-NOV-1995;   95US-0552142.
XX
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA   (WEIN/) WEINTRAUB N.
XX
PI   Hollenberg SM,  Lee JE,  Tapscott SJ,  Weintraub HM;
XX
DR   WPI; 1997-272117/24.
DR   P-PSDB; AAW22439.
XX
PT   Nucleic acid encoding neurogenic differentiation polypeptide -
PT   useful e.g. in regulating neuronal, endocrine and gastrointestinal
PT   development
XX
PS   Claim 1; Page 61-62; 81pp; English.
XX
CC   Neurogenic differentiation (NeuroD) genes (AAT74887-94) and proteins
CC   (AAW22436-43) from human, mouse and frog have been identified,
CC   isolated and sequence.  NeuroD polypeptides are tissue-specific
CC   basic-helix-loop-helix (bHLH) transcriptional activators involved
CC   in neuronal, endocrine and gastrointestinal development.  They were
CC   discovered by expression cloning and screening assays designed to
CC   identify possible bHLH proteins capable of interacting with the
CC   protein product of the Drosophila daugherless gene.  Novel neuroD2
CC   and neuroD3 genes, related to neuroD1, have been identified.
CC   NeuroD nucleic acids can be used to produce NeuroD polypeptides,
CC   construction of test cell lines, as probes, in gene therapy, and to
CC   produce transgenic animals as models of disease.
XX
SQ   Sequence 1535 BP; 250 A; 559 C; 476 G; 244 T; 6 other;

  Query Match             1.4%;  Score 21;  DB 18;  Length 1535;
  Best Local Similarity   100.0%;  Pred. No. 6.4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        729 CGGGAGCGCAACCGCATGCAC 749
              |||||||||||||||||||||
Db        439 CGGGAGCGCAACCGCATGCAC 459


RESULT 35
AAV42931
ID   AAV42931 standard; DNA; 1535 BP.
XX
AC   AAV42931;
XX
DT   25-MAR-2003  (updated)
DT   21-OCT-1998  (first entry)
XX
DE   DNA encoding human neuroD2 protein, which is a bHLH protein.
XX
KW   Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;
KW   classification; medulloblastoma; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             55..1194
FT                   /*tag=  a
FT                   /product= neuroD2
XX
PN   US5795723-A.
XX
PD   18-AUG-1998.
XX
PF   07-AUG-1997;   97US-0910973.
XX
PR   06-MAY-1994;   94US-0239238.
PR   02-NOV-1995;   95US-0552142.
PR   30-OCT-1996;   96WO-US17532.
XX
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI   Olson JM,  Tapscott SJ;
XX
DR   WPI; 1998-466661/40.
DR   P-PSDB; AAW71015.
XX
PT   Classifying neuroectodermal tumours from expression pattern of
PT   basic-helix-loop-helix genes - especially for identifying
```

```
PT   medulla:blastoma and assessing its aggressiveness, specifically
PT   associated with expression of BHLH genes neuroD 1-3
XX
PS   Example 11; Columns 57-60; 43pp; English.
XX
CC   The present sequence encodes a protein which is a member of the basic
CC   helix-loop-helix (bHLH) protein family, and is designated neuroD2. The
CC   bHLH genes are a family of genes associated with vertebrate neuronal,
CC   endocrinal and gastrointestinal development. The observed pattern of
CC   neuroD expression distinguishes subclasses of neuroectodermal tumours.
CC   The specification describes a method for the classification of human
CC   neuroectodermal tumours. The method comprises measuring, in a tumour
CC   sample, expression of at least one basic bHLH gene and identifying the
CC   tumour subclass by matching expression to predetermined expression
CC   profiles for known subclasses. For classifying the tumour as a
CC   medulloblastoma, the bHLH gene detected is neuroD1 and neuroD3.
CC   The method is used to classify neuroectodermal tumours, and to identify
CC   medullablastoma and for prognosis of this as aggressive.
CC   (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ   Sequence 1535 BP; 250 A; 559 C; 476 G; 244 T; 6 other;

  Query Match             1.4%;  Score 21;  DB 19;  Length 1535;
  Best Local Similarity   100.0%;  Pred. No. 6.4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        729 CGGGAGCGCAACCGCATGCAC 749
              |||||||||||||||||||||
Db        439 CGGGAGCGCAACCGCATGCAC 459


RESULT 36
ABS56389
ID   ABS56389 standard; DNA; 1535 BP.
XX
AC   ABS56389;
XX
DT   23-JAN-2003  (first entry)
XX
DE   Human bHLH family neuroD2 genomic DNA, clone 14B1.
XX
KW   Human; gene; ds; neuroD3; neuroD; basic-helix-loop-helix; bHLH;
KW   differentiation; neurone; endocrine; gastrointestine; development;
KW   transgenic; embryo; birth defect; spontaneous abortion; stem cell;
KW   cancer; neural growth factor; tumour; diagnostic; motor; sensory;
KW   traumatic neural injury; hearing; vision; brain; spinal cord;
KW   malabsorption syndrome; gastrointestinal dysmotility syndrome;
KW   Hirsh Prung's disease; therapeutic; neuroD2.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             55..1194
FT                   /*tag=  a
FT                   /product= "NeuroD2"
FT   misc_feature    463..582
FT                   /*tag=  b
FT                   /note= "HLH coding domain"
XX
PN   US6444463-B1.
XX
PD   03-SEP-2002.
XX
PF   07-FEB-2000; 2000US-0499227.
XX
PR   05-AUG-1998;   98WO-US16417.
PR   06-MAY-1994;   94US-0239238.
PR   08-MAY-1995;   95WO-US05741.
PR   02-NOV-1995;   95US-0552142.
PR   30-OCT-1996;   96WO-US17532.
PR   07-AUG-1997;   97US-0910973.
XX
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI   Tapscott SJ;
XX
DR   WPI; 2003-056678/05.
DR   P-PSDB; ABG72001.
XX
PT   New neurogenic differentiation gene, useful in gene therapy to correct
PT   traumatic neural injury that has resulted in loss of motor or sensory
PT   neural function and for constructing recombinant cell lines  -
XX
PS   Example 11; Column 57-60; 43pp; English.
XX
CC   The invention discloses an isolated nucleic acid molecule which encodes a
CC   functionally active human neuroD3 polypeptide. NeuroD proteins represent
CC   a new family within the basic-helix-loop-helix (bHLH) family which are
CC   implicated in the regulation of differentiation. NeuroD proteins are
CC   particularly involved in neuronal, endocrine and gastrointestinal
CC   development. The nucleic acid is useful for constructing recombinant cell
CC   lines, transgenic embryos and animals and for quantifying the level of
CC   expression of neuroD in a cell. Birth defects and spontaneous abortions
CC   may result from expression of an abnormal neuroD protein. The
CC   polynucleotide sequences permit the establishment of primary cultures of
CC   proliferating embryonic neuronal stem cells under conditions mimicking
CC   those that are active in development and cancer. The resultant cell lines
CC   find use as sources of novel neural growth factors, in assays for
CC   identifying novel neuronal growth factors which can be used for screening
CC   anti-cancer drugs capable of driving terminal differentiation in neural
CC   tumours, for producing antibodies useful in diagnostic assays and for
CC   screening for compounds capable of modulating the activity of neuroD.
CC   Transformed host cells, nucleic acids and polypeptides are also useful
CC   for treating sites of traumatic neural injury where motor or sensory
CC   neural activity has been lost, e.g. hearing or vision loss and brain or
CC   spinal cord damage. The host cells find use in the treatment of
CC   malabsorption syndromes or gastrointestinal dysmotility syndromes (Hirsh
CC   Prung's Disease). The cell lines also find use in screening for candidate
CC   therapeutic agents capable of either substituting for neuroD or
CC   correcting the cellular defect caused by a defective neuroD. The sequence
CC   presented is the human neuroD2 genomic DNA, clone 14B1.
XX
SQ   Sequence 1535 BP; 250 A; 559 C; 476 G; 244 T; 6 other;

  Query Match             1.4%;  Score 21;  DB 25;  Length 1535;
  Best Local Similarity   100.0%;  Pred. No. 6.4;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        729 CGGGAGCGCAACCGCATGCAC 749
              |||||||||||||||||||||
Db        439 CGGGAGCGCAACCGCATGCAC 459


RESULT 37
AAA62681
ID   AAA62681 standard; DNA; 1550 BP.
XX
AC   AAA62681;
XX
DT   29-NOV-2000  (first entry)
XX
DE   Human NeuroD2 gene.
XX
KW   Human; NeuroD2; neurogenic basic helix-loop-helix protein;
KW   epidermal cell transdifferentiation; gene therapy; cerebroprotective;
KW   neuroprotective; brain injury; spinal cord injury; stroke;
KW   neurodegenerative disease; Parkinson's disease; Huntington's disease;
KW   Alzheimer's disease; neuronal cell generation; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             58..1206
FT                   /*tag=  a
FT                   /product= "NeuroD2 protein"
XX
```

PN US6087168-A.
XX
PD 11-JUL-2000.
XX
PF 20-JAN-1999; 99US-0234332.
XX
PR 20-JAN-1999; 99US-0234332.
XX
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
PI Levesque MF, Neuman T;
XX
DR WPI; 2000-498200/44.
DR P-PSDB; AAB14347.
XX
PT Converting epidermal cells into neurons, useful for isolating nerve
PT growth factors or for gene therapy, comprises dedifferentiating cells
PT and transfecting with vectors with a cDNA coding neurogenic
PT transcription factors -
XX
PS Example 2; Column 19-22; 27pp; English.
XX
CC The present sequence is the human neurogenic basic helix-loop-helix
CC protein (NeuroD2) gene from Genbank. It was used to provide sequence
CC information for the cloning of NeuroD2 cDNA, which was used to
CC transfect cultured epidermal cells. This was part of a novel method
CC for transdifferentiating an epidermal basal cell into a cell having the
CC morphological, physiological and/or immunological features of a viable
CC neuronal cell. The method is useful for screening new drugs for
CC treating a nervous system disorder, or for isolating a novel nerve
CC growth factor. The transdifferentiated cell is useful in both cell and
CC gene therapies aimed at alleviating various neurological disorders. The
CC cell or gene therapy approach involves the use of autologous
CC transplantation or grafting of the newly created neuronal cells as
CC treatment for brain or spinal cord injury, stroke and neurodegenerative
CC diseases (e.g. Parkinson's disease, Huntington's disease or Alzheimer's
CC disease).
XX
SQ Sequence 1550 BP; 250 A; 569 C; 486 G; 243 T; 2 other;

Query Match 1.4%; Score 21; DB 21; Length 1550;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 729 CGGGAGCGCAACCGCATGCAC 749
||||||||||||||||||||
Db 442 CGGGAGCGCAACCGCATGCAC 462

RESULT 38
AAD46888
ID AAD46888 standard; DNA; 1665 BP.
XX
AC AAD46888;
XX
DT 27-JAN-2003 (first entry)
XX
DE Human neurogenin 1 (Ngn1) gene #1.
XX
KW Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW islet cell; cell therapy; neurogenin 1; Ngn1; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 260..973
FT /*tag= a
FT /product= "Human Ngn1 protein"
XX
PN WO200274045-A2.
XX
PD 26-SEP-2002.
XX
PF 20-MAR-2002; 2002WO-US11166.
XX
PR 20-MAR-2001; 2001US-0817360.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI German MS, Lin J;
XX
DR WPI; 2002-759853/82.
DR P-PSDB; AAE29279.
XX
PT Producing a mammalian islet cell for treating diabetes mellitus
PT comprises introducing into a mammalian cell a nucleic acid molecule
PT encoding neuroendocrine basic helix-loop-helix transcription factor -
XX
PS Disclosure; Page 93; 108pp; English.
XX
CC The invention relates to a method for producing a mammalian islet cell.
CC The method comprising introducing into a mammalian cell a nucleic acid
CC molecule encoding an islet transcription factor for expression of the
CC islet transcription factor in the cell and for production of islet cell
CC phenotype in the cell. The islet transcription factor is a neuroendocrine
CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC for treating type 2 diabetes mellitus and for replacing beta cells lost
CC to autoimmune destruction in individuals with type 1 diabetes. The method
CC is useful in cell therapy. The present sequence is human neurogenin 1
CC (Ngn1) gene.
XX
SQ Sequence 1665 BP; 344 A; 561 C; 443 G; 317 T; 0 other;

Query Match 1.4%; Score 21; DB 24; Length 1665;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAAC 752
||||||||||||||||||||
Db 560 GAGCGCAACCGCATGCACAAC 580

RESULT 39
AAL04043/c
ID AAL04043 standard; DNA; 2776 BP.
XX
AC AAL04043;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 6731.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 6731; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 2776 BP; 631 A; 851 C; 749 G; 545 T; 0 other;

Query Match 1.4%; Score 21; DB 22; Length 2776;
Best Local Similarity 100.0%; Pred. No. 6.3;

```
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1112 AACAGGCCCTGGGCGGTGGGC 1132
              |||||||||||||||||||||
Db        938 AACAGGCCCTGGGCGGTGGGC 918


RESULT 40
AAL04045/c
ID   AAL04045 standard; DNA; 2776 BP.
XX
AC   AAL04045;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6733.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
```

PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 6733; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 2776 BP; 631 A; 853 C; 747 G; 545 T; 0 other;

Query Match 1.4%; Score 21; DB 22; Length 2776;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1112 AACAGGCCCTGGGCGGTGGGC 1132
||||||||||||||||||||
Db 938 AACAGGCCCTGGGCGGTGGGC 918

RESULT 41
AAK68475
ID AAK68475 standard; DNA; 2776 BP.
XX
AC AAK68475;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23287.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 23287; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 2776 BP; 545 A; 749 C; 851 G; 631 T; 0 other;

Query Match 1.4%; Score 21; DB 22; Length 2776;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1112 AACAGGCCCTGGGCGGTGGGC 1132
||||||||||||||||||||
Db 1839 AACAGGCCCTGGGCGGTGGGC 1859

RESULT 42
AAK68476
ID AAK68476 standard; DNA; 2776 BP.
XX
AC AAK68476;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23288.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 23288; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 2776 BP; 545 A; 747 C; 853 G; 631 T; 0 other;

Query Match 1.4%; Score 21; DB 22; Length 2776;
Best Local Similarity 100.0%; Pred. No. 6.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1112 AACAGGCCCTGGGCGGTGGGC 1132
              |||||||||||||||||||||
Db       1839 AACAGGCCCTGGGCGGTGGGC 1859
```

RESULT 43
ABS69045/c
ID ABS69045 standard; DNA; 352 BP.
XX
AC ABS69045;
XX
DT 21-NOV-2002 (first entry)
XX
DE Novel murine polynucleotide isolated using gene trap technology #108.
XX
KW Mouse; gene trapped sequence; GTS; functional genomic analysis;
KW phage display system; gene chip; temporal gene expression;
KW tissue specific gene expression; antisense inhibition; gene targeting;
KW development disorder; cell differentiation disorder; aging; cancer;
KW autoimmune disease; lupus; inflammatory disorder; skin disorder;
KW degenerative disorder; ds.
XX
OS Mus musculus.
XX
PN US2002102543-A1.
XX
PD 01-AUG-2002.
XX
PF 30-NOV-2000; 2000US-0728445.
XX
PR 01-DEC-1999; 99US-168358P.
XX
PA (FRIE/) FRIEDRICH G.
PA (ZAMB/) ZAMBROWICZ B.
PA (SAND/) SANDS A T.
XX
PI Friedrich G, Zambrowicz B, Sands AT;
XX
DR WPI; 2002-690598/74.
XX
PT Novel murine polynucleotides that individually identify novel genes
PT into which a retroviral gene trap vector has integrated, useful in
PT genomic analysis and in discovery, development of therapeutic and
PT diagnostic agents -
XX
PS Claim 1; Page 57; 296pp; English.
XX
CC The invention describes an isolated murine polynucleotide (I) comprising
CC a contiguous stretch of at least 60 nucleotides of one of 265-677
CC nucleotide 891 OMNIBANK gene trapped sequences (GTSs) (S), given in the
CC specification. The novel genes and cells are useful in functional
CC genomic analysis and in the discovery and development of new therapeutic
CC and diagnostic agents and methods. (I) is useful for identifying the
CC coding regions of the murine genome, to isolate cDNAs, genomic clones,
CC or full-length genes/polynucleotides or homologues, heterologues,
CC paralogues, or orthologues that are capable of hybridising to one or more
CC of the GTSs under stringent conditions. (I) can be incorporated into a
CC phage display system that can be used to screen for proteins, or other
CC ligands, that are capable of binding an amino acid sequence encoded by
CC an oligonucleotide or polynucleotide sequence in at least one of the TS
CC sequences. (I) is useful in addressable arrays, such as gene chips, to
CC identify and characterise temporal and tissue specific gene expression,
CC to identify the gene of interest from many sources and for genetic
CC manipulations such as antisense inhibition and gene targeting. Decreasing
CC the level of expression of (I) and/or down regulating the activity of
CC peptides or proteins encoded by (I) is useful for treating development
CC and cell differentiation disorders, aging, cancer, autoimmune disease,
CC lupus, inflammatory disorders, skin disorders and degenerative
CC disorders. This sequence represents a murine cDNA isolated using gene
CC trap technology.
XX
SQ Sequence 352 BP; 97 A; 89 C; 84 G; 82 T; 0 other;

Query Match 1.4%; Score 20; DB 24; Length 352;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1010 CACAGCCTCATTGGAGGAGT 1029
              ||||||||||||||||||||
Db        209 CACAGCCTCATTGGAGGAGT 190
```

RESULT 44
ABQ49522/c
ID ABQ49522 standard; DNA; 592 BP.
XX
AC ABQ49522;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36113.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 592 BP; 81 A; 59 C; 201 G; 251 T; 0 other;

Query Match 1.4%; Score 20; DB 24; Length 592;

Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        834 CGCTTCGCCCACAACTACAT 853
              ||||||||||||||||||||
Db         68 CGCTTCGCCCACAACTACAT 49
```

RESULT 45
ABQ49523
ID ABQ49523 standard; DNA; 592 BP.
XX
AC ABQ49523;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36114.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;

Query Match 1.4%; Score 20; DB 24; Length 592;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        834 CGCTTCGCCCACAACTACAT 853
              ||||||||||||||||||||
Db        525 CGCTTCGCCCACAACTACAT 544
```

Search completed: January 26, 2004, 22:22:39
Job time : 428 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 19:54:07 ; Search time 565 Seconds
(without alignments)
9417.792 Million cell updates/sec

Title: US-09-595-947E-1
Perfect score: 1460
Sequence: 1 gcaggtagcgagaggagcag.....agagtgacctaatccagtgt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2434939 seqs, 1822278265 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	92	6.3	645	14	US-10-004-717-4	Sequence 4, Appli
	2	92	6.3	861	14	US-10-004-717-24	Sequence 24, Appl
	3	92	6.3	1861	9	US-09-817-360-3	Sequence 3, Appli
	4	36	2.5	65	13	US-09-908-975-4140	Sequence 4140, Ap
	5	35	2.4	65	13	US-09-908-975-30269	Sequence 30269, A
	6	32	2.2	5340	9	US-09-817-360-1	Sequence 1, Appli
c	7	26	1.8	26	11	US-09-992-665-221	Sequence 221, App
	8	23	1.6	738	8	US-08-722-570-13	Sequence 13, Appl
	9	23	1.6	790	14	US-10-004-717-20	Sequence 20, Appl
	10	23	1.6	1385	14	US-10-004-717-30	Sequence 30, Appl
	11	23	1.6	1412	14	US-10-004-717-6	Sequence 6, Appli
	12	23	1.6	1412	14	US-10-004-717-37	Sequence 37, Appl
	13	21	1.4	21	15	US-10-090-011-1	Sequence 1, Appli
	14	21	1.4	21	15	US-10-090-011-49	Sequence 49, Appl
	15	21	1.4	576	13	US-10-029-386-8222	Sequence 8222, Ap
c	16	21	1.4	723	13	US-10-027-632-129767	Sequence 129767,
c	17	21	1.4	723	13	US-10-027-632-129768	Sequence 129768,
c	18	21	1.4	723	14	US-10-027-632-129767	Sequence 129767,
c	19	21	1.4	723	14	US-10-027-632-129768	Sequence 129768,
c	20	21	1.4	2776	11	US-09-764-891-6731	Sequence 6731, Ap
c	21	21	1.4	2776	11	US-09-764-891-6733	Sequence 6733, Ap
c	22	20	1.4	352	10	US-09-728-445-108	Sequence 108, App
c	23	19	1.3	473	11	US-09-918-995-33698	Sequence 33698, A
c	24	19	1.3	500	10	US-09-783-590-2157	Sequence 2157, Ap
c	25	19	1.3	501	12	US-10-260-238-5886	Sequence 5886, Ap
	26	19	1.3	582	13	US-10-027-632-276355	Sequence 276355,
	27	19	1.3	582	14	US-10-027-632-276355	Sequence 276355,
	28	19	1.3	735	13	US-10-027-632-157510	Sequence 157510,
	29	19	1.3	735	14	US-10-027-632-157510	Sequence 157510,
c	30	19	1.3	1263	10	US-09-938-842A-1036	Sequence 1036, Ap
c	31	19	1.3	1263	12	US-09-938-842A-1036	Sequence 1036, Ap
	32	19	1.3	2180	13	US-10-027-632-111282	Sequence 111282,
	33	19	1.3	2180	13	US-10-027-632-111283	Sequence 111283,
	34	19	1.3	2180	13	US-10-027-632-111284	Sequence 111284,
	35	19	1.3	2180	14	US-10-027-632-111282	Sequence 111282,
	36	19	1.3	2180	14	US-10-027-632-111283	Sequence 111283,
	37	19	1.3	2180	14	US-10-027-632-111284	Sequence 111284,
c	38	19	1.3	3424	11	US-09-764-891-8887	Sequence 8887, Ap
c	39	19	1.3	3425	11	US-09-764-891-8888	Sequence 8888, Ap
	40	18	1.2	169	12	US-10-242-535A-3008	Sequence 3008, Ap
	41	18	1.2	466	13	US-10-027-632-184196	Sequence 184196,
	42	18	1.2	466	14	US-10-027-632-184196	Sequence 184196,
	43	18	1.2	502	13	US-10-029-386-2219	Sequence 2219, Ap
c	44	18	1.2	504	13	US-10-029-386-6504	Sequence 6504, Ap
	45	18	1.2	637	13	US-10-027-632-208140	Sequence 208140,

ALIGNMENTS

RESULT 1
US-10-004-717-4
; Sequence 4, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-4

Query Match 6.3%; Score 92; DB 14; Length 645;
Best Local Similarity 100.0%; Pred. No. 3.9e-37;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        304 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 363

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||
Db        364 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 395
```

```
RESULT 2
US-10-004-717-24
; Sequence 24, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 861
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-10-004-717-24

  Query Match             6.3%;  Score 92;  DB 14;  Length 861;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-37;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 522

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||
Db        523 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 554


RESULT 3
US-09-817-360-3
; Sequence 3, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
;  APPLICANT: German, Michael S.
;  APPLICANT:  Lin, Joseph
;  TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
;  TITLE OF INVENTION:  AND DELIVERY OF INSULIN
;  FILE REFERENCE: UCSF-129CIP
;  CURRENT APPLICATION NUMBER: US/09/817,360
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: 09/535,145
;  PRIOR FILING DATE: 2000-03-24
;  PRIOR APPLICATION NUMBER: 60/128,180
;  PRIOR FILING DATE: 1999-04-06
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1861
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-817-360-3

  Query Match             6.3%;  Score 92;  DB 9;  Length 1861;
  Best Local Similarity   100.0%;  Pred. No. 3.5e-37;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1396 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 1455

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              |||||||||||||||||||||||||||||||
Db       1456 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 1487


RESULT 4
US-09-908-975-4140
; Sequence 4140, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4140
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-908-975-4140

  Query Match             2.5%;  Score 36;  DB 13;  Length 65;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-08;
  Matches   36;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1281 ATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGT 1316
              ||||||||||||||||||||||||||||||||||||
Db         30 ATTGCAGGCTGATCTCCTCTTAACCCTCCTCAGTGT 65


RESULT 5
US-09-908-975-30269
; Sequence 30269, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
;  APPLICANT: SHOSHAN, Avi
;  APPLICANT:  WASSERMAN, Alon
;  APPLICANT:  MINTZ, Eli
;  APPLICANT:  MINTZ, Liat
;  APPLICANT:  FAIGLER, Simchon
;  TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
;  TITLE OF INVENTION:  THAT POPULATE A TRANSCRIPTOME
;  FILE REFERENCE: 36688-0005
;  CURRENT APPLICATION NUMBER: US/09/908,975
;  CURRENT FILING DATE:  2001-07-20
;  PRIOR APPLICATION NUMBER: US 60/287,724
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: US 60/221,607
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 32337
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30269
;   LENGTH: 65
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-908-975-30269

  Query Match             2.4%;  Score 35;  DB 13;  Length 65;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-07;
  Matches   35;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        547 TCAGTTCCAATTCCACCCCACCTAGCCCCACTCTC 581
              |||||||||||||||||||||||||||||||||||
Db          1 TCAGTTCCAATTCCACCCCACCTAGCCCCACTCTC 35


RESULT 6
US-09-817-360-1
; Sequence 1, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
;  APPLICANT: German, Michael S.
;  APPLICANT:  Lin, Joseph
;  TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
;  TITLE OF INVENTION:  AND DELIVERY OF INSULIN
;  FILE REFERENCE: UCSF-129CIP
;  CURRENT APPLICATION NUMBER: US/09/817,360
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: 09/535,145
;  PRIOR FILING DATE: 2000-03-24
;  PRIOR APPLICATION NUMBER: 60/128,180
;  PRIOR FILING DATE: 1999-04-06
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 5340
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
US-09-817-360-1

  Query Match             2.2%;  Score 32;  DB 9;  Length 5340;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-06;
  Matches   32;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        672 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 703
              ||||||||||||||||||||||||||||||||
Db       3235 AAGAGCGAGTTGGCACTGAGCAAGCAGCGACG 3266


RESULT 7
US-09-992-665-221/c
; Sequence 221, Application US/09992665
; Publication No. US20030092009A1
; GENERAL INFORMATION:
;  APPLICANT: Kaia Palm
;  TITLE OF INVENTION: PROFILING TUMOR SPECIFIC MARKERS FOR THE
;  TITLE OF INVENTION:  DIAGNOSIS AND TREATMENT OF NEOPLASTIC DISEASE
;  FILE REFERENCE: CEMINES.002A
;  CURRENT APPLICATION NUMBER: US/09/992,665
;  CURRENT FILING DATE:  2001-11-13
;  PRIOR APPLICATION NUMBER: 60/249,508
;  PRIOR FILING DATE: 2000-11-16
;  NUMBER OF SEQ ID NOS: 380
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 221
;   LENGTH: 26
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Probe
US-09-992-665-221

  Query Match             1.8%;  Score 26;  DB 11;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 0.01;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        673 AGAGCGAGTTGGCACTGAGCAAGCAG 698
              ||||||||||||||||||||||||||
Db         26 AGAGCGAGTTGGCACTGAGCAAGCAG 1


RESULT 8
US-08-722-570-13
; Sequence 13, Application US/08722570
; Publication No. US20030044887A1
;  GENERAL INFORMATION:
;    APPLICANT:  Anderson, David J.
;    APPLICANT:  Ma, Qiufu
;    TITLE OF INVENTION:  NEUROGENIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,570
;      FILING DATE:  27-SEP-1996
;      CLASSIFICATION:  5365
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-63902/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  738 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-08-722-570-13

  Query Match             1.6%;  Score 23;  DB 8;  Length 738;
  Best Local Similarity   100.0%;  Pred. No. 0.26;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        304 GAGCGCAACCGCATGCACAACCT 326


RESULT 9
US-10-004-717-20
; Sequence 20, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
;  APPLICANT: ZOGHBI, HUDA Y.
;  APPLICANT:  YANG, QI
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
;  TITLE OF INVENTION:  ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
;  TITLE OF INVENTION:  OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
;  FILE REFERENCE: P01899US4
;  CURRENT APPLICATION NUMBER: US/10/004,717
;  CURRENT FILING DATE:  2002-08-16
;  PRIOR APPLICATION NUMBER: 09/585,645
;  PRIOR FILING DATE: 2000-06-01
;  PRIOR APPLICATION NUMBER: 60/176,993
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: 60/137,060
;  PRIOR FILING DATE: 1999-06-01
;  NUMBER OF SEQ ID NOS: 69
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
;   LENGTH: 790
;   TYPE: DNA
;   ORGANISM: chicken
```

US-10-004-717-20

Query Match 1.6%; Score 23; DB 14; Length 790;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACCT 754
||||||||||||||||||||||
Db 374 GAGCGCAACCGCATGCACAACCT 396

RESULT 10
US-10-004-717-30
; Sequence 30, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-30

Query Match 1.6%; Score 23; DB 14; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACCT 754
||||||||||||||||||||||
Db 742 GAGCGCAACCGCATGCACAACCT 764

RESULT 11
US-10-004-717-6
; Sequence 6, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-6

Query Match 1.6%; Score 23; DB 14; Length 1412;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACCT 754
||||||||||||||||||||||
Db 445 GAGCGCAACCGCATGCACAACCT 467

RESULT 12
US-10-004-717-37
; Sequence 37, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-37

Query Match 1.6%; Score 23; DB 14; Length 1412;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACCT 754
||||||||||||||||||||||
Db 445 GAGCGCAACCGCATGCACAACCT 467

RESULT 13
US-10-090-011-1
; Sequence 1, Application US/10090011
; Publication No. US20030082810A1
; GENERAL INFORMATION:
; APPLICANT: Serup, Palle
; APPLICANT: Heimberg, Harry
; APPLICANT: Gradwohl Gerard
; TITLE OF INVENTION: Methods For Generating Insulin-Secreting
; TITLE OF INVENTION: Cells Suitable for Transplantation
; FILE REFERENCE: 6246.200-US
; CURRENT APPLICATION NUMBER: US/10/090,011
; CURRENT FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: US 60/271,474
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-090-011-1

Query Match 1.4%; Score 21; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        460 TGGCGCCTCATCCCTTGGATG 480
              |||||||||||||||||||||
Db          1 TGGCGCCTCATCCCTTGGATG 21


RESULT 14
US-10-090-011-49
; Sequence 49, Application US/10090011
; Publication No. US20030082810A1
; GENERAL INFORMATION:
;  APPLICANT: Serup, Palle
;  APPLICANT:  Heimberg, Harry
;  APPLICANT:  Gradwohl Gerard
;  TITLE OF INVENTION: Methods For Generating Insulin-Secreting
;  TITLE OF INVENTION:  Cells Suitable for Transplantation
;  FILE REFERENCE: 6246.200-US
;  CURRENT APPLICATION NUMBER: US/10/090,011
;  CURRENT FILING DATE:  2002-02-26
;  PRIOR APPLICATION NUMBER: US 60/271,474
;  PRIOR FILING DATE: 2001-02-26
;  NUMBER OF SEQ ID NOS: 70
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49
;   LENGTH: 21
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-090-011-49

  Query Match             1.4%;  Score 21;  DB 15;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 4.1;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        798 CCGGATGACGCCAAACTTACA 818
              |||||||||||||||||||||
Db          1 CCGGATGACGCCAAACTTACA 21


RESULT 15
US-10-029-386-8222
; Sequence 8222, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;  TITLE OF INVENTION:  EXPRESSION ANALYSIS TWO
;  FILE REFERENCE: AEOMICA-X-2
;  CURRENT APPLICATION NUMBER: US/10/029,386
;  CURRENT FILING DATE:  2001-12-20
;  NUMBER OF SEQ ID NOS: 34288
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 8222
;   LENGTH: 576
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO CHR16.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  67
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =  64
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.8
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  41
;   OTHER INFORMATION: EST_HUMAN HIT: BG166801.1, EVALUE 0.00e+00
;   OTHER INFORMATION: NT HIT: gi14779902, EVALUE 3.00e-95
;   OTHER INFORMATION: SWISSPROT HIT: P25444, EVALUE 8.00e-58
US-10-029-386-8222

  Query Match             1.4%;  Score 21;  DB 13;  Length 576;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1112 AACAGGCCCTGGGCGGTGGGC 1132
              |||||||||||||||||||||
Db         29 AACAGGCCCTGGGCGGTGGGC 49


Search completed: January 29, 2004, 22:30:14
Job time : 568 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 19:42:42 ; Search time 2899 Seconds
(without alignments)
12240.267 Million cell updates/sec

Title: US-09-595-947E-1
Perfect score: 1460
Sequence: 1 gcaggtagcgagaggagcag.....agagtgacctaatccagtgt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	92	6.3	593	28	AZ296526	AZ296526 RPCI-23-1
	2	60	4.1	1025	14	BY708009	BY708009 BY708009
	3	60	4.1	1540	11	AK008017	AK008017 Mus muscu
	4	24	1.6	923	13	BX445903	BX445903 BX445903
	5	23	1.6	504	4	BX524720	Bx524720 RZPD Mus
	6	23	1.6	517	13	BU053975	BU053975 UI-M-FD0-
	7	23	1.6	600	12	BG808248	BG808248 2083-52 M
	8	23	1.6	600	13	BU924937	BU924937 7103-91 M
	9	23	1.6	673	13	BU709606	BU709606 UI-M-FR0-
	10	23	1.6	687	14	CA945402	CA945402 UI-M-FD0-
	11	23	1.6	689	13	BU058877	BU058877 UI-M-FR0-
	12	23	1.6	730	13	BU612495	BU612495 UI-M-FR0-
	13	23	1.6	814	13	BQ178789	BQ178789 UI-M-EV0-
	14	23	1.6	823	13	BU054481	BU054481 UI-M-FD0-
	15	23	1.6	932	14	CA979119	CA979119 AGENCOURT
	16	23	1.6	1001	9	AL540071	AL540071 AL540071
	17	23	1.6	1269	12	BG854922	BG854922 1024041C0
	18	22	1.5	589	13	BU775885	BU775885 SJEBRD03
	19	21	1.4	319	10	BE936551	BE936551 RC5-NT005
	20	21	1.4	464	28	AQ753599	AQ753599 HS_2117_A
	21	21	1.4	536	28	AZ506899	AZ506899 1M0348G13
	22	21	1.4	704	13	BU057851	BU057851 UI-M-FR0-
	23	21	1.4	718	10	BE263765	BE263765 601194122
	24	21	1.4	740	14	CA319439	CA319439 UI-M-FW0-
	25	21	1.4	781	13	BU611678	BU611678 UI-M-FI0-
	26	21	1.4	829	12	BI910102	BI910102 603067946
	27	21	1.4	835	14	CA320553	CA320553 UI-M-FW0-
	28	21	1.4	913	10	BF204175	BF204175 601867625
	29	21	1.4	953	9	AU067624	AU067624 AU067624
	30	21	1.4	962	13	BQ686909	BQ686909 AGENCOURT
	31	21	1.4	1022	12	BM563662	BM563662 AGENCOURT
	32	21	1.4	1039	10	BE780690	BE780690 601469349
	33	21	1.4	1042	10	BG419220	BG419220 602445870
	34	20	1.4	241	9	AA113743	AA113743 W5F Pyroc
c	35	20	1.4	351	14	T70947	T70947 yc49c05.r1
	36	20	1.4	397	12	BI445873	BI445873 dai33c04.
c	37	20	1.4	444	9	AV944801	AV944801 AV944801
c	38	20	1.4	476	12	BM253695	BM253695 514849 MA
	39	20	1.4	550	9	AV939670	AV939670 AV939670
	40	20	1.4	909	14	CD325295	CD325295 AGENCOURT
c	41	20	1.4	971	29	CNS01YSD	AL173398 Tetraodon
	42	20	1.4	971	29	CNS02YD5	AL219506 Tetraodon
c	43	20	1.4	1010	13	BX377288	BX377288 BX377288
	44	20	1.4	1075	12	BM810917	BM810917 AGENCOURT
	45	19	1.3	201	10	BB071548	BB071548 BB071548

ALIGNMENTS

RESULT 1
AZ296526/c
LOCUS AZ296526 593 bp DNA linear GSS 27-JUL-2000
DEFINITION RPCI-23-160G18.TV RPCI-23 Mus musculus genomic clone RPCI-23-160G18 , genomic survey sequence.
ACCESSION AZ296526
VERSION AZ296526.1 GI:9538311
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 593)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished
COMMENT Other_GSSs: RPCI-23-160G18.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 160 row: G column: 18
Seq primer: T7
Class: BAC ends.

```
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-160G18"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_lib="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and partially digested
                     with a combination of EcoRI and EcoRI Methylase. Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI sites. The ligation products were transformed into
                     DH10B electrocompetent cells (BRL Life Technologies). "
BASE COUNT       88 a    159 c    213 g    133 t
ORIGIN

  Query Match             6.3%;  Score 92;  DB 28;  Length 593;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-34;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 191

Qy        822 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 853
              ||||||||||||||||||||||||||||||||
Db        190 ATCGAGACCCTGCGCTTCGCCCACAACTACAT 159
```

RESULT 2
BY708009

```
LOCUS        BY708009                1025 bp    mRNA    linear   EST 16-DEC-2002
DEFINITION   BY708009 RIKEN full-length enriched, adult male small intestine Mus
             musculus cDNA clone 2010001M19 5', mRNA sequence.
ACCESSION    BY708009
VERSION      BY708009.1  GI:27119192
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 1025)
  AUTHORS    Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
             Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
             , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
             Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
             Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
             Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
             ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
             ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
             , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
             Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
             King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
             ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
             ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
             Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
             Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
             ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
             ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
             ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
             Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
             ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
             Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
             ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
             , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
             ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
             ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
             ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE      Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
  JOURNAL    Nature 420, 563-573 (2002)
  MEDLINE    22354683
   PUBMED    12466851
COMMENT      Contact: Yoshihide Hayashizaki
             Laboratory for Genome Exploration Research Group, RIKEN Genomic
             Sciences Center(GSC), Yokohama Institute
             The Institute of Physical and Chemical Research (RIKEN)
             1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             Tel: 81-45-503-9222
             Fax: 81-45-503-9216
             Email: genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/
             Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
             ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
             Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
             ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
             Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
             Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
             Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
             Submission
              Computational Analysis of Full-Length Mouse cDNAs Compared with
             Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
              Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new
             genes. Genome Res. 10 (10), 1617-1630 (2000)
              RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer. Genome Res.
             10 (11), 1757-1771 (2000)
              Computer-based methods for the mouse full-length cDNA
             encyclopedia: real-time sequence clustering for construction of a
             nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
              cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
              Please visit our web site (http://genome.gsc.riken.go.jp) for
             further details.
FEATURES             Location/Qualifiers
     source          1. .1025
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="2010001M19"
                     /sex="male"
                     /tissue_type="small intestine"
                     /dev_stage="adult"
                     /lab_host="SOLR"
                     /clone_lib="RIKEN full-length enriched, adult male small
                     intestine"
                     /note="Site_1: XhoI; Site_2: SstI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 5.0 and subtraction to Rot = 20.0. Second strand
                     cDNA was prepared with the primer adapter of sequence [5'
                     GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC
```

3']. cDNA was cleaved with XhoI and SstI."
BASE COUNT 215 a 316 c 286 g 207 t 1 others
ORIGIN

Query Match 4.1%; Score 60; DB 14; Length 1025;
Best Local Similarity 100.0%; Pred. No. 1.4e-18;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 544 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 603

RESULT 3
AK008017
LOCUS AK008017 1540 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010001M19 product:atonal homolog 5 (Drosophila), full insert sequence.
ACCESSION AK008017
VERSION AK008017.1 GI:12841941
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1540)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.
FEATURES Location/Qualifiers
source 1. .1540
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:2010001M19"
/db_xref="MGI:1907403"
/db_xref="taxon:10090"
/clone="2010001M19"
/sex="male"
/tissue_type="small intestine"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 241. .885
/note="unnamed protein product; atonal homolog 5 (Drosophila) (MGD|MGI:893591, GB|NM_009719, evidence: BLASTN, 99%, match=648)
putative"
/codon_start=1
/protein_id="BAB25411.1"
/db_xref="GI:12841942"
/db_xref="MGI:893591"
/translation="MAPHPLDALTIQVSPETQQPFPGASDHEVLSSNSTPPSHTLIPR
DCSEAEVGDCRGTSRKLRARRGGRNRPKSELALSKKRRSRRKKANDRERNRMHNLNSA
LDALRGVLPTFPDDAKLTKVETLRFAHNYIWALTQTLRIADHSLYGPEPPVPCGELGS
PGGGSNGDWGSIYSPVSQAGNLSPTASLEEFPGLQVPSSPSYLLPGALVFSDFL"

```
BASE COUNT      306 a    485 c    397 g    352 t
ORIGIN

  Query Match             4.1%;  Score 60;  DB 11;  Length 1540;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-18;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        762 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        544 GCGCTGGATGCGCTGCGCGGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAG 603


RESULT 4
BX445903
LOCUS       BX445903                923 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX445903 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF015YD06 5-PRIME, mRNA sequence.
ACCESSION   BX445903
VERSION     BX445903.1  GI:30774371
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2626.r
            Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS2BAX8ZA08_AX14ZE4_1.
FEATURES             Location/Qualifiers
     source          1. .923
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF015YD06"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
BASE COUNT      167 a    290 c    214 g    251 t      1 others
ORIGIN

  Query Match             1.6%;  Score 24;  DB 13;  Length 923;
  Best Local Similarity   100.0%;  Pred. No. 1.6;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        820 AGATCGAGACCCTGCGCTTCGCCC 843
              ||||||||||||||||||||||||
Db         18 AGATCGAGACCCTGCGCTTCGCCC 41


RESULT 5
BX524720
ID   BX524720   standard; RNA; EST; 504 BP.
XX
AC   BX524720;
XX
SV   BX524720.1
XX
DT   27-MAY-2003 (Rel. 75, Created)
DT   27-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   RZPD Mus musculus cDNA clone IMAGp998O187204 = IMAGE:2922473 5' EST.
XX
KW   EST; expressed sequence tag.
XX
OS   Mus musculus (house mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX
RN   [1]
RP   1-504
RA   Heil O., Ebert L., Neubert P., Peters M., Radelof U., Schneider D.,
RA   Korn B.;
RT   ;
RL   Submitted (28-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Im Neuenheimer
RL   Feld 580, D-69120 Heidelberg, Germany
XX
CC   RZPD; IMAGp998O187204.
CC   RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
CC   Mouse UnigeneSet - RZPD2 (RZPDLIB No.981)
CC   http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi/response?libNo=981
CC   Contact: Ina Rolfs
CC   RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
CC   Heubnerweg 6, D-14059 Berlin, Germany
CC   Tel: +49 30 32639 101
CC   Fax: +49 30 32639 111
CC   www.rzpd.de
CC   This clone is available royalty-free from RZPD;
CC   contact RZPD (clone@rzpd.de) for further information.
CC   Seq primer: T7, Primer sequence: TAATACGACTCACTATAGGG
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .504
FT                   /db_xref="taxon:10090"
FT                   /note="Cloned unidirectionally. Primer: Oligo dT.
FT                   Size-selected 1.5 kb for average insert size 2 kb. Primary
FT                   library; non-amplified. This library was constructed by M.
FT                   Mione (University College London, Dept of Anatomy and
FT                   Developmental Biology)."
FT                   /organism="Mus musculus"
FT                   /clone="IMAGp998O187204"
FT                   /clone_lib="Mione mouse WTB"
FT                   /dev_stage="13.5 dpc"
FT                   /lab_host="DH12S"
XX
SQ   Sequence 504 BP; 90 A; 162 C; 192 G; 60 T; 0 other;

  Query Match             1.6%;  Score 23;  DB 4;  Length 504;
  Best Local Similarity   100.0%;  Pred. No. 4.7;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        471 GAGCGCAACCGCATGCACAACCT 493


RESULT 6
BU053975
LOCUS       BU053975                517 bp    mRNA    linear   EST 26-AUG-2002
DEFINITION  UI-M-FD0-bzg-h-11-0-UI.r1 NIH_BMAP_FD0 Mus musculus cDNA clone
            IMAGE:6403258 5', mRNA sequence.
ACCESSION   BU053975
VERSION     BU053975.1  GI:22494052
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
```

```
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6403258"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FD0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     , is TGAGAGAGCC. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instututes of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT       90 a    166 c    194 g     67 t
ORIGIN

  Query Match             1.6%;  Score 23;  DB 13;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 4.8;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        449 GAGCGCAACCGCATGCACAACCT 471


RESULT 7
BG808248
LOCUS       BG808248                600 bp    mRNA    linear   EST 20-DEC-2001
DEFINITION  2083-52 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
            mRNA sequence.
ACCESSION   BG808248
VERSION     BG808248.1  GI:17955225
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
            White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
  TITLE     Gene expression in the developing mouse retina by EST sequencing
            and microarray analysis
  JOURNAL   Nucleic Acids Res. 29 (24), 4983-4993 (2001)
  MEDLINE   21671825
   PUBMED   11812828
COMMENT     Contact: Klein WH
            Department of Biochemistry and Molecular Biology
            University of Texas M.D. Anderson Cancer Center
            Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
            Tel: 713 792 3646
            Fax: 713 790 0329.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="neural retina"
                     /dev_stage="embryonic day 14.5 post-fertilization"
                     /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
BASE COUNT       94 a    238 c    161 g    107 t
ORIGIN

  Query Match             1.6%;  Score 23;  DB 12;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 4.9;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db          4 GAGCGCAACCGCATGCACAACCT 26


RESULT 8
BU924937
LOCUS       BU924937                600 bp    mRNA    linear   EST 30-OCT-2002
DEFINITION  7103-91 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
            mRNA sequence.
ACCESSION   BU924937
VERSION     BU924937.1  GI:24428820
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
            White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
  TITLE     Gene expression in the developing mouse retina by EST sequencing
            and microarray analysis
  JOURNAL   Nucleic Acids Res. 29 (24), 4983-4993 (2001)
  MEDLINE   21671825
   PUBMED   11812828
COMMENT     Contact: Klein WH
            Department of Biochemistry and Molecular Biology
            University of Texas M.D. Anderson Cancer Center
            Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
            Tel: 713 792 3646
            Fax: 713 790 0329.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="neural retina"
                     /dev_stage="embryonic day 14.5 post-fertilization"
                     /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
BASE COUNT       99 a    194 c    222 g     82 t      3 others
ORIGIN

  Query Match             1.6%;  Score 23;  DB 13;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 4.9;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        363 GAGCGCAACCGCATGCACAACCT 385


RESULT 9
BU709606
LOCUS       BU709606                673 bp    mRNA    linear   EST 26-NOV-2002
```

```
DEFINITION  UI-M-FR0-cbe-g-01-0-UI.r1 NIH_BMAP_FR0 Mus musculus cDNA clone
            IMAGE: 6808442 5', mRNA sequence.
ACCESSION   BU709606
VERSION     BU709606.1  GI:23643245
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6808442"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FR0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT      122 a    219 c    250 g     81 t      1 others
ORIGIN

  Query Match             1.6%;  Score 23;  DB 13;  Length 673;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        570 GAGCGCAACCGCATGCACAACCT 592


RESULT 10
CA945402
LOCUS       CA945402                687 bp    mRNA    linear   EST 30-DEC-2002
DEFINITION  UI-M-FD0-cdh-1-12-0-UI.r1 NIH_BMAP_FD0 Mus musculus cDNA clone
            IMAGE: 6828925 5', mRNA sequence.
ACCESSION   CA945402
VERSION     CA945402.1  GI:27433882
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 687)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .687
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6828925"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FD0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     , is TGAGAGAGCC. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instututes of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      117 a    228 c    250 g     91 t      1 others
ORIGIN

  Query Match             1.6%;  Score 23;  DB 14;  Length 687;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        449 GAGCGCAACCGCATGCACAACCT 471


RESULT 11
BU058877
LOCUS       BU058877                689 bp    mRNA    linear   EST 26-AUG-2002
DEFINITION  UI-M-FR0-cak-k-23-0-UI.r1 NIH_BMAP_FR0 Mus musculus cDNA clone
            IMAGE:6413710 5', mRNA sequence.
ACCESSION   BU058877
VERSION     BU058877.1  GI:22499166
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES              Location/Qualifiers
     source           1. .689
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:6413710"
                      /tissue_type="whole brain"
                      /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                      /lab_host="DH10B (T1 phage resistant)"
                      /clone_lib="NIH_BMAP_FR0"
                      /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                      Site_2: Not I; The library was constructed according
                      Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                      1996. Denatured RNA was size fractionated on a 1% agarose
                      gel. First strand cDNA synthesis was primed with oligo-dT
                      primer containing a Not I site. Double strand cDNA was
                      size selected according to mRNA size fraction, ligated
                      with EcoR I adaptor, digested with NotI and then cloned
                      directionally into pYX-Asc vector. The library tag
                      sequence located between the Not I site and the polyA tail
                      is AGCGAGACAG. This library was created for the University
                      Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                      Developing Mouse Nervous System', supported by National
                      Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                      program coordinator."
BASE COUNT      128 a    228 c    255 g     78 t
ORIGIN

  Query Match             1.6%;  Score 23;  DB 13;  Length 689;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        626 GAGCGCAACCGCATGCACAACCT 648


RESULT 12
BU612495
LOCUS       BU612495                730 bp    mRNA    linear   EST 20-FEB-2003
DEFINITION  UI-M-FR0-cbc-k-21-0-UI.r1 NIH_BMAP_FR0 Mus musculus cDNA clone
            UI-M-FR0-cbc-k-21-0-UI 5', mRNA sequence.
ACCESSION   BU612495
VERSION     BU612495.1  GI:23278710
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 730)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES              Location/Qualifiers
     source           1. .730
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6"
                      /db_xref="taxon:10090"
                      /clone="UI-M-FR0-cbc-k-21-0-UI"
                      /tissue_type="whole brain"
                      /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                      /lab_host="DH10B (T1 phage resistant)"
                      /clone_lib="NIH_BMAP_FR0"
                      /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                      Site_2: Not I; The library was constructed according
                      Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                      1996. Denatured RNA was size fractionated on a 1% agarose
                      gel. First strand cDNA synthesis was primed with oligo-dT
                      primer containing a Not I site. Double strand cDNA was
                      size selected according to mRNA size fraction, ligated
                      with EcoR I adaptor, digested with NotI and then cloned
                      directionally into pYX-Asc vector. The library tag
                      sequence located between the Not I site and the polyA tail
                      is AGCGAGACAG. This library was created for the University
                      Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                      Developing Mouse Nervous System', supported by National
                      Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                      program coordinator."
BASE COUNT      131 a    244 c    264 g     91 t
ORIGIN

  Query Match             1.6%;  Score 23;  DB 13;  Length 730;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        570 GAGCGCAACCGCATGCACAACCT 592


RESULT 13
BQ178789
LOCUS       BQ178789                814 bp    mRNA    linear   EST 30-APR-2002
DEFINITION  UI-M-EV0-bwt-k-07-0-UI.r1 NIH_BMAP_EV0 Mus musculus cDNA clone
            IMAGE:5701758 5', mRNA sequence.
ACCESSION   BQ178789
VERSION     BQ178789.1  GI:20354281
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 814)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES              Location/Qualifiers
     source           1. .814
                      /organism="Mus musculus"
                      /mol_type="mRNA"
```

```
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5701758"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 15.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_EV0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site.  Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     , is GTGCGTGGAA. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instututes of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      151 a    268 c    291 g    102 t      2 others
ORIGIN

  Query Match             1.6%;  Score 23;  DB 13;  Length 814;
  Best Local Similarity   100.0%;  Pred. No. 5.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        663 GAGCGCAACCGCATGCACAACCT 685


RESULT 14
BU054481
LOCUS       BU054481                823 bp    mRNA    linear   EST 26-AUG-2002
DEFINITION  UI-M-FD0-bzj-i-24-0-UI.r1 NIH_BMAP_FD0 Mus musculus cDNA clone
            IMAGE:6404447 5', mRNA sequence.
ACCESSION   BU054481
VERSION     BU054481.1  GI:22494558
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 823)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .823
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6404447"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 12.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FD0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNa was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     , is TGAGAGAGCC. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instututes of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      155 a    273 c    289 g    104 t      2 others
ORIGIN

  Query Match             1.6%;  Score 23;  DB 13;  Length 823;
  Best Local Similarity   100.0%;  Pred. No. 5.1;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        672 GAGCGCAACCGCATGCACAACCT 694


RESULT 15
CA979119
LOCUS       CA979119                932 bp    mRNA    linear   EST 06-JAN-2003
DEFINITION  AGENCOURT_11295215 NIH_MGC_164 Mus musculus cDNA clone
            IMAGE:30146192 5', mRNA sequence.
ACCESSION   CA979119
VERSION     CA979119.1  GI:27511773
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe and Dr. Mina
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM0061  row: h  column: 09
            High quality sequence start: 16
            High quality sequence stop: 640.
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30146192"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_164"
                     /note="Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2:
                     NotI; Non-normalized full-length enriched library from
                     pooled mouse embryonic limb, maxilla and mandible, day
                     10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
                     Cloned directionally, priming method: Oligo-dT.  cDNA
                     enrichment: >1k bp, Average insert size 1.8k bp. Priming
                     sequence: 5'GACTAGTTCTAGATCGCGAGCGGCCGCCC(T) 3'. Tissue
                     contributed by, David Rowe. Library constructed by ResGen,
                     Invitrogen Corp."
BASE COUNT      153 a    349 c    276 g    153 t      1 others
```

ORIGIN

Query Match 1.6%; Score 23; DB 14; Length 932;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        732 GAGCGCAACCGCATGCACAACCT 754
              |||||||||||||||||||||||
Db        561 GAGCGCAACCGCATGCACAACCT 583
```

Search completed: January 29, 2004, 22:18:31
Job time : 2908 secs